(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 330 052**

(51) Int. Cl.:
C12N 15/10 (2006.01)
C12N 15/62 (2006.01)
G01N 33/50 (2006.01)
G01N 33/68 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **07010609 .1**
(96) Fecha de presentación : **27.02.1992**
(97) Número de publicación de la solicitud: **1820858**
(97) Fecha de publicación de la solicitud: **22.08.2007**

(54) Título: **Proteína quimérica que comprende micro-proteínas que tienen dos o más puentes disulfuro y realizaciones de las mismas.**

(30) Prioridad: **01.03.1991 US 664989**

(45) Fecha de publicación de la mención BOPI:
03.12.2009

(45) Fecha de la publicación del folleto de la patente:
03.12.2009

(73) Titular/es: **Dyax Corporation**
300 Technology Square, 8th Floor
Cambridge, Massachusetts 02139, US

(72) Inventor/es: **Ladner, Robert Charles;**
Roberts, Bruce Lindsay;
Ley, Arthur Charles y
Kent, Rachel Baribault

(74) Agente: **Ungría López, Javier**

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).

Venta de fascículos: Oficina Española de Patentes y Marcas. Pº de la Castellana, 75 – 28071 Madrid

DESCRIPCIÓN

Proteína quimérica que comprende micro-proteínas que tienen dos o más puentes disulfuro y realizaciones de las mismas.

Antecedentes de la invención

Campo de la invención

Esta invención se refiere a una biblioteca de proteínas quiméricas, comprendiendo cada proteína quimérica (a) una micro-proteína de entre seis y cuarenta aminoácidos y (b) una secuencia de aminoácidos obtenida de una proteína de la superficie externa de un paquete genético, donde dicha proteína quimérica se presenta en la superficie externa de dicho paquete genético; un proceso para identificar proteínas con una actividad de unión deseada contra una diana, comprendiendo el proceso explorar la biblioteca; una proteína quimérica expresada por la biblioteca; y una mezcla de ácidos nucleicos que codifican la biblioteca.

Descripción de la técnica relacionada

La secuencia de aminoácidos de una proteína determina su estructura tridimensional (3D), que a su vez determina la función de la proteína. Algunos restos de la cadena polipeptídica son más importantes que otros para determinar la estructura 3D de una proteína, y por lo tanto su capacidad para unirse, de forma no covalente, pero muy fuerte y específicamente, a moléculas diana características.

La "modificación genética de proteínas" es la técnica de manipular la secuencia de una proteína para, por ejemplo, alterar sus características de unión. Los factores que afectan a la unión de la proteína se conocen, pero el diseño de nuevas superficies complementarias ha demostrado ser difícil. Quiocho *et. al*. (QUIO87) sugieren que es improbable que, usando los actuales métodos de modificación genética de proteínas, puedan construirse proteínas con propiedades de unión superiores a las de las proteínas de origen natural.

Sin embargo, se han producido algunos éxitos aislados. Por ejemplo, Wilkinson *et al*. (WILK84) informaron de que un mutante de la tirosil ARNt sintetasa de *Bacillus stearothermophilus* con la mutación Thr_{51}→Pro muestra un aumento de 100 veces en la afinidad por ATP.

Con el desarrollo de técnicas de ADN recombinante, se hizo posible obtener una proteína mutante mutando el gen que codifica la proteína nativa y después expresando el gen mutado. Se conocen varias estrategias de mutagénesis. Una, la "cirugía de la proteína" (DILL87), implica la introducción de una o más mutaciones *predeterminadas* dentro de gen de elección. Se expresa un *único* polipéptido de secuencia totalmente predeterminada y se evalúan sus características de unión.

En el otro extremo está la mutagénesis aleatoria por medio de mutágenos relativamente no específicos tales como radiación y diversos agentes químicos. Véase Ho *et al*. (HOCJ85) y Lehtovaara, Solicitud EP 285.123.

Es posible modificar aleatoriamente nucleótidos predeterminados usando una mezcla de bases en los ciclos apropiados de un procedimiento de síntesis de ácido nucleico. (OLIP86, OLIP87) La proporción de bases en la mezcla, para cada posición de un codón, determinará la frecuencia con la cual cada aminoácido se producirá en los polipéptidos expresados a partir de la población de ADN degenerado. (REID88a; VERS86a; VERS86b). El problema de la abundancia desigual de ADN que codifica diferentes aminoácidos no se analiza.

Ferenci y colaboradores han publicado una serie de documentos sobre el aislamiento cromatográfico de mutantes de la proteína del transporte de maltosa LambB de *E. coli* (FERE82a, FERE82b, FERE83, FERE84, CLUN84, HEIN87 y documentos mencionados en esos documentos). Los mutantes eran espontáneos o inducidos con mutágenos químicos no específicos. Los niveles de mutación se tomaron para proporcionar mutaciones puntuales únicas o inserciones únicas de dos restos. No se buscaron ni se descubrieron mutaciones múltiples.

Aunque se observó variación en el grado de afinidad por los sustratos de LamB convencionales maltosa y almidón, no había selección por afinidad a una molécula diana a la que no se unía en absoluto LamB nativo, y no se buscaron o descubrieron mutaciones múltiples. FERE84 suponía que la técnica de selección cromatográfica por afinidad podía adaptarse para desarrollar mutantes similares de otras "importantes enzimas situadas en la superficie bacteriana" y para seleccionar mutaciones que dieran como resultado la recolocación de una proteína bacteriana intracelular en la superficie celular. Las proteínas superficiales mutantes de Ferenci no podían ser, sin embargo, quimeras de una proteína de superficie bacteriana y un dominio de unión exógeno o heterólogo.

Ferenci también demostró que no había necesidad de clonar el gen estructural o de conocer la estructura, el sitio activo o la secuencia de la proteína. El método de la presente invención, sin embargo, utiliza específicamente un gen estructural clonado. No es posible construir un gen que codifique una proteína de unión potencial, quimérica, dirigida a la superficie externa sin clonación.

Ferenci no limitaba las mutaciones a loci particulares. Las sustituciones estaban limitadas por la naturaleza del mutágeno en lugar de por la conveniencia de un tipo de aminoácido particular en un sitio particular. En la presente invención, se usa el conocimiento de la estructura, sitio activo y/o secuencia de la proteína para predecir qué restos son más probables que afecten a la actividad de unión sin desestabilizar de forma indebida la proteína y la mutagénesis se centra en esos sitios. Ferenci no sugiere que deban modificarse preferiblemente restos superficiales. Como consecuencia, el sistema de selección de Ferenci es mucho menos eficaz que el que se describe en este documento.

Varios investigadores han dirigido epítopos antigénicos extraños *no mutados* a la superficie de bacterias o de fagos, fusionados con una proteína superficial bacteriana o del fago nativo y demostraron que los epítopos eran reconocidos por anticuerpos. De este modo, Charbit, *et al*. (CHAR86a,b) insertaron genéticamente el epítopo C3 de la proteína de cubierta VP1 del poliovirus en la proteína de la membrana externa LamB de *E. coli* y determinaron inmunológicamente que el epítopo C3 estaba expuesto en la superficie celular bacteriana. Charbit, *et al*. (CHAR87) produjeron del mismo modo quimeras de LamB y los epítopos A (o B) de la región preS2 del virus de la hepatitis B.

Una proteína quimérica LacZ/OmpB se ha expresado en *E. coli* y se dirige, dependiendo de la fusión, a la membrana externa o al periplasma (SILH77). También se ha expresado una proteína superficial LacZ/OmpA quimérica y se ha presentado en la superficie de células de *E. coli* (WEIN83). Otros investigadores han expresado y presentado en la superficie de una célula quimeras de otras proteínas superficiales bacterianas, tales como fimbrias de tipo 1 de *E. coli* (HEDE89) y fimbrias de tipo 1 de *Bacterioides nodusus* (JENN89). En ninguno de los casos mencionados se mutagenizó el material genético insertado.

Dulbecco (DULB86) sugiere un procedimiento para incorporar un epítopo antigénico extraño en una proteína de superficie viral, de modo que la proteína quimérica expresada se presente en la superficie del virus de tal manera que el epítopo extraño quede accesible a un anticuerpo. En 1985 Smith (SMIT85) presentó la inserción de un segmento no funcional del gen de la endonucleasa *Eco*RI en el gen *III* del bacteriófago fl, "en fase". La proteína del gen III es una proteína de cobertura secundaria necesaria para la infectividad. Smith demostró que el fago recombinante se adsorbía por un anticuerpo inmovilizado generado contra la endonucleasa *Eco*RI y podía eluirse con ácido. De la Cruz *et al*. (DELA88) han expresado un fragmento de la región repetida de la proteína del circunsporozoíto de *Plasmodium falciparum* en la superficie de M13 en forma de inserto en la proteína del gen III. Demostraron que el fago recombinante era antigénico e inmunogénico en conejos, y que dicho fago recombinante podía usarse para el mapeado del epítopo B. Los investigadores sugieren que puede usarse un fago recombinante similar para el mapeado del epítopo T y para el desarrollo de vacunas.

Ninguno de estos investigadores sugirió la mutagénesis del material insertado, ni el material insertado es un dominio de unión completo que otorga a la proteína quimérica la capacidad de unirse específicamente a un receptor diferente del sitio de combinación de un antígeno con un anticuerpo.

McCafferty *et al*. (MCCA90) expresaron una fusión de un fragmento Fv de un anticuerpo al N-terminal de la proteína pIII. El fragmento Fv no se mutó.

Parmley y Smith (PARM88) sugirieron que podría construirse y usarse una biblioteca de epítopos que mostrase todos los hexapéptidos posibles para aislar epítopos que se unan a anticuerpos. Al analizar la biblioteca de epítopos, los autores no sugirieron que era deseable equilibrar la representación de diferentes aminoácidos. Ni tampoco muestran que el inserto debe codificar un dominio completo de la proteína exógena. Los epítopos se consideran péptidos no estructurados en oposición a las proteínas estructuradas.

Scott y Smith (SCOT90) y Cwirla *et al*. (CWIR90) prepararon "bibliotecas de epítopos" en las que los epítopos hexapéptidos potenciales para un anticuerpo diana se mutaron aleatoriamente fusionando oligonucleótidos degenerados, que codifican los epítopos, con el gen III del fago fd y expresando el gen fusionado en células infectadas con el fago. Las células fabricaron fago de fusión que presentaba los epítopos en su superficie; el fago que se unía al anticuerpo inmovilizado se eluía con ácido y se analizaba. En ambos casos, el gen fusionado se caracterizaba por un segmento que codificaba una región espaciadora para separar la región variable de la secuencia pIII de tipo silvestre de modo que los aminoácidos modificados no estarían restringidos por la secuencia pIII cercana. Devlin *et al*. (DEVL90) exploraron análogamente, usando un fago M13, 15 restos de epítopos reconocidos por la estreptavidina. De nuevo, se usó un espaciador para separar los péptidos aleatorios del resto de la proteína quimérica del fago. Por lo tanto, estas referencias están lejos de mostrar la restricción del repertorio conformacional de los restos mutados.

Otro problema con las bibliotecas de Scott y Smitt, Cwirla *et al*. y Devlin *et al*. era que proporcionaban un muestreo muy sesgado de los posibles aminoácidos en cada posición. Su primera preocupación a la hora de diseñar el oligonucleótido degenerado que codificaba su región variable era asegurarse de que los veinte aminoácidos eran codificables en cada posición; una segunda preocupación era minimizar la frecuencia de aparición de señales de detención. Como consecuencia, Scott y Smith y Cwirla *et al*. emplearon NNK (N = mezcla equivalente de G, A, T, C; K = mezcla equivalente de G y T) mientras que Devlin *et al*. usaron NNS (S = mezcla equivalente de G y C). No hicieron intentos de minimizar la proporción de frecuencia entre el aminoácido más favorecido al menos favorecido o de igualar el índice de aparición de aminoácidos ácidos y básicos.

Devlin *et al*. caracterizaron varios péptidos de unión a estreptavidina seleccionados por afinidad, pero no midieron las constantes de afinidad de estos péptidos. Cwirla *et al*. determinaron las constantes de afinidad para estos péptidos,

pero quedaron decepcionados al descubrir que sus mejores hexapéptidos tenían afinidades (350-300 nM), "órdenes de magnitud" más débiles que las del epítopo Met-encefalina (7 nM) reconocido por el anticuerpo diana. Cwirla *et al*. suponían que los péptidos que tenían el fago con afinidades mayores permanecían unidos en elución ácida, posiblemente por causa de las interacciones multivalentes entre el fago (que tenía aproximadamente 4 copias de pIII) y la IgG diana divalente. Scott y Smith fueron capaces de descubrir péptidos cuya afinidad por el anticuerpo diana (A2) era comparable a la del epítopo miohemeritrina de referencia (50 nM). Sin embargo, Scott y Smith expresaron también su preocupación porque algunos péptidos de alta afinidad se perdían, posiblemente a través de la unión irreversible del fago de fusión a la diana.

Lam, *et al*. (LAM91) crearon una biblioteca pentapeptídica mediante síntesis no biológica en soportes sólidos. Aunque muestran que es deseable obtener todos los pentapéptidos aleatorios posibles en proporciones aproximadamente equimolares, excluyeron deliberadamente la cisteína para eliminar cualquier posibilidad de reticulación por disulfuro.

Ladner, Glick y Bird, documento WO 88/06630 (publicado el 7 de septiembre de 1988 y que tiene prioridad de la solicitud de Estados Unidos 07/021.046, asignada a Genex Corp.) (LGB) suponen que pueden seleccionarse diversos dominios de anticuerpo de cadena sencilla (SCAD) para la unión a un antígeno particular modificando el ADN que codifica las regiones determinantes combinadas de un anticuerpo de cadena sencilla, subclonando el gen SCAD en el gen gpV del fago λ de modo que se disponga una quimera SCAD/gpV en la superficie externa del fago λ, y seleccionando el fago que se une al antígeno a través de cromatografía de afinidad. El único antígeno mencionado es la hormona del crecimiento bovina. No se analizan otras moléculas de unión, dianas, organismos transportadores o proteínas de la superficie externa. Tampoco se hace mención al método o grado de mutagénesis. Además, tampoco se muestra la estructura exacta de la fusión o se analiza cómo identificar una fusión con éxito o cómo proceder si el SCAD no se muestra.

Ladner y Bird, documento WO 88/06601 (publicado el 7 de septiembre de 1988) sugieren que pueden prepararse represores "pseudodiméricos" de cadena sencilla (proteínas de unión a ADN) mutando un péptido de unión supuesto seguido por la selección *in vivo* de esta mutación y la selección puede usarse para crear un diccionario de elementos de reconocimiento para su uso en el diseño de represores asimétricos. Los represores no se presentan en la superficie externa de un organismo.

Los métodos para identificar restos en una proteína que pueden sustituirse con una cisteína para promover la formación de un puente disulfuro que estabiliza la proteína se proporcionan en los documentos Pantoliano y Ladner, Patente de Estados Unidos Nº 4.903.773 (PANT90), Pantoliano y Ladner (PANT87), Pabo y Suchenek (PABO86), MATS89 y SAUE86.

Ladner, *et al*., documento WO 90/02809 describe mutagénesis semialeatorias ("variegación") de proteínas conocidas que se muestran como dominios de proteínas de la superficie externa semiartificiales de bacterias, fagos o esporas, y la selección por afinidad de mutantes que tienen características de unión deseadas. Las proteínas más pequeñas mencionadas específicamente en el documento WO 90/02809 son crambina (disulfuros 3:40, 4:32, 16:26; 46 aminoácidos), el tercer dominio del ovomucoide (disulfuros 8:38, 16:35 y 24:56, 56 aminoácidos) y BPTI (disulfuros 5:55, 14:38, 30:51; 58 aminoácidos). El documento WO 90/02809 también describe específicamente una estrategia para "variegar" un codón para obtener una mezcla de los veinte aminoácidos en esa posición en proporciones aproximadamente iguales.

Bass, *et al*. (BASS90) fusionaron hormona del crecimiento humana con la proteína del gen III del fago M13. Sugiere que pueden mutarse hGH y otras "grandes proteínas" y pueden aplicarse "selecciones de unión".

Sumario de la invención

Un polipéptido es un polímero compuesto por una única cadena del mismo o de diferentes aminoácidos unidos por enlaces peptídicos. Los péptidos lineales pueden tomar una gran cantidad de conformaciones diferentes a través de giros internos alrededor de los enlaces sencillos de la cadena principal de cada carbono. Estos giros están limitados a grados variables por los grupos laterales, siendo la glicina la que menos limita y la valina, isoleucina y especialmente la prolina las que más. Un polipéptido de 20 restos puede tener 10^{20} conformaciones diferentes que puede asumir mediante diversas rotaciones internas.

Las proteínas son polipéptidos que, como resultado de interacciones de estabilización entre aminoácidos que no están necesariamente en posiciones adyacentes en la cadena, se han plegado en una conformación bien definida. Este plegamiento es normalmente esencial para su actividad biológica.

Para polipéptidos de 40-60 restos o más, las fuerzas no covalentes, tales como puentes de hidrógeno, puentes salinos e interacciones hidrófobas son suficientes para estabilizar un plegamiento o conformación particular. Los segmentos constituyentes del polipéptido se mantienen más o menos en esta conformación a menos que se vea perturbada por un agente desnaturalizante tal como alta temperatura o un pH bajo o alto, con lo cual el polipéptido se despliega o se "funde". Cuanto más pequeño sea el péptido, más probable es que su conformación esté determinada por el ambiente. Si un péptido pequeño no restringido tiene actividad biológica, el ligando peptídico será esencialmente un enrollamiento aleatorio hasta que se acerca al receptor. El receptor acepta al péptido solamente en una o unas pocas

conformaciones porque las conformaciones alternativas están desfavorecidas por fuerzas de van der Waals y otras interacciones no covalentes no favorables.

Los polipéptidos pequeños tienen ventajas potenciales sobre los polipéptidos mayores cuando se usan como agentes terapéuticos o de diagnóstico, incluyendo (aunque sin limitación):

a) mejor penetración en los tejidos,

b) eliminación más rápida de la circulación (importante para agentes formadores de imágenes),

c) menor antigenicidad, y

d) mayor actividad por masa.

Además, los polipéptidos, especialmente los de menos de 40 restos aproximadamente, tienen la ventaja de la accesibilidad mediante síntesis química; se prefieren particularmente polipéptidos de menos de aproximadamente 30 restos. De este modo, sería deseable ser capaz de emplear la combinación de mutación y selección por afinidad para identificar pequeños polipéptidos que se unan a una diana de elección.

La mayoría de los polipéptidos de este tamaño, sin embargo, tienen desventajas como moléculas de unión. De acuerdo con Olivera *et al*. (OLIV90a): "Los péptidos en este intervalo de tamaño se equilibran normalmente entre muchas conformaciones (para tener una conformación fija, las proteínas tienen que ser generalmente mucho mayores)". La unión específica de un péptido a una molécula diana requiere que el péptido tome una conformación que sea complementaria al sitio de unión. Para un decapéptido con tres conformaciones isoenergéticas (por ejemplo, lámina β, α hélice y giro inverso) en cada resto, hay aproximadamente $6{\cdot}10^4$ conformaciones generales posibles. Suponiendo que estas conformaciones son equiprobables para el decapéptido no limitado, si solamente una de las posibles conformaciones se une al sitio de unión, entonces la afinidad del péptido por la diana se espera que sea aproximadamente $6{\cdot}10^4$ más alta si este péptido pudiera limitarse a la única conformación eficaz. De este modo, podría esperarse que el decapéptido no limitado, en relación con el decapéptido limitado a la conformación correcta, muestre una afinidad inferior. También podría mostrar una especificidad inferior, puesto que una de las otras conformaciones del decapéptido no limitado puede ser una que se una fuertemente a un material diferente de la diana pretendida. Como corolario, podría tener menos resistencia a la degradación por proteasas, puesto que podía ser más probable que proporcione un sitio de unión para la proteasa.

La presente invención supera estos problemas, mientras que conserva las ventajas de los polipéptidos más pequeños, identificando proteínas quiméricas nuevas que tienen las características de unión deseadas. Las mini-proteínas son pequeños polipéptidos que, aunque son demasiado pequeños para tener una conformación estable como resultado de fuerzas no covalentes en solitario, están reticuladas covalentemente (por ejemplo, mediante puentes disulfuro) en una conformación estable y de este modo tienen actividades biológicas más típicas de moléculas proteicas mayores que de los polipéptidos no limitados de tamaño comparable. Las mini-proteínas a las que se refiere la presente invención son micro-proteínas unidas por puentes disulfuro de entre seis a 40 aminoácidos, teniendo la micro-proteína dos o más enlaces disulfuro, en el que cada enlace disulfuro está formado por un par de cisteínas invariables.

La presente invención se refiere a la construcción, expresión y selección de genes mutados específicos para nuevas proteínas quiméricas con propiedades de unión deseables así como a las propias proteínas quiméricas y a las "bibliotecas" de "paquetes genéticos" mutantes usados para presentar las proteínas quiméricas a un material "diana" potencial. Las "dianas" pueden ser, aunque no es necesario, proteínas. Las dianas pueden incluir otras macromoléculas biológicas o sintéticas así como otras sustancias orgánicas o inorgánicas.

La solicitud anterior, WO 90/02809 muestra generalmente que pueden mutarse dominios proteicos estables para identificar nuevas proteínas con características de unión deseables. Entre las proteínas "parentales" adecuadas que se identifican específicamente como útiles para este fin están tres proteínas - BPTI (58 restos), el tercer dominio del ovomucoide (56 restos) y crambina (46 restos) - que están en el intervalo de tamaño de 40-60 restos en las que las interacciones no covalentes entre aminoácidos no adyacentes se vuelven significativas; las tres contienen también tres enlaces disulfuro que potencian la estabilidad de la molécula.

En ninguna parte del documento WO 90/02809 se encuentra ningún reconocimiento específico de que un polipéptido con menos de 40 restos, y especialmente aquéllos con solamente uno o dos enlaces disulfuro, tendrían la suficiente estabilidad para servir como "armazón" para una variación mutacional. Sin embargo, estas "micro-proteínas" son de gran utilidad como se ha indicado anteriormente.

El documento WO 90/02809 sugiere también el uso de una proteína, azurina, que tiene una forma de reticulación diferente (Cu:CYS,HIS,HIS,MET). Sin embargo, la azurina tiene 128 aminoácidos, por lo tanto no puede posiblemente considerarse una mini-proteína. La presente invención se refiere al uso de mini-proteínas de menos de 60 aminoácidos que presentan una reticulación coordinada por un ión metálico.

Gracias a la presente invención, se obtienen proteínas que pueden unirse específicamente a dianas diferentes de los sitios de combinación del antígeno a los anticuerpos. No debe considerarse a una proteína como "proteína de unión"

solamente porque pueda unirse a un anticuerpo (véase la definición de "proteína de unión" que se muestra a continuación). Aunque casi cualquier secuencia de aminoácidos de más de aproximadamente 6-8 aminoácidos es probable que, cuando se una a un vehículo inmunogénico, provoque una respuesta inmune, es improbable que cualquier polipéptido aleatorio dado satisfaga la rigurosa definición de "proteína de unión" con respecto a la afinidad mínima y a la especificidad por su sustrato. Solamente ensayando muchos polipéptidos aleatorios simultáneamente (y, en el caso habitual, controlando el grado y el carácter de la variación de la secuencia, es decir, limitándola a restos de un dominio de unión potencial que tiene una estructura estable, seleccionándose los restos como los más probables que afecten a la unión y a la estabilidad), se supera este obstáculo.

Las reivindicaciones adjuntas se incorporan por la presente como referencia en esta memoria descriptiva como una lista de las realizaciones preferidas.

Breve descripción de las figuras

La Figura 1 muestra la cadena principal de toxina de escorpión (Brookhaven Protein Data Bank entrada 1SN3) restos 20 al 42. CYS_{25} y CYS_{41} se muestran formando un puente disulfuro. En la proteína nativa, estos grupos forman puentes disulfuro con otras cisteínas, pero no se requiere movimiento de la cadena principal para poner los azufres gamma en geometría aceptable. Los restos diferentes de GLY están marcados en el carbono β con el código de una letra.

Descripción detallada de las realizaciones preferidas

I. *Introducción*

El principio fundamental de la invención es el de la *evolución forzada*. En la naturaleza, la evolución resulta de la combinación de la variación genética, selección de rasgos ventajosos y reproducción de los individuos seleccionados, enriqueciendo de este modo la población con el rasgo. La presente invención consigue variación genética a través de mutagénesis aleatoria controlada ("variegación") del ADN, produciendo una mezcla de moléculas de ADN que codifican dominios de unión potenciales diferentes pero relacionados que son mutantes de micro-proteínas. La invención selecciona genes mutados específicos para nuevas proteínas con propiedades de unión deseables 1) disponiendo que el producto de cada gen mutado se *presente* en la superficie externa de un paquete genético replicable (GP) (una célula, espora o virus) que contenga el gen y 2) usando selección por *selección por afinidad* para unir al material diana - para enriquecer la población de paquetes en aquellos paquetes que contienen genes que especifican proteínas con unión mejorada a este material diana. Finalmente, se consigue el enriquecimiento permitiendo que se reproduzcan solamente los paquetes genéticos que gracias a la proteína desplegada, se unen a la diana. La evolución se "fuerza" en esta selección por el material diana proporcionado y porque los codones particulares se mutagenizan a frecuencias mayores que la natural.

La estrategia de *presentación* se perfecciona en primer lugar modificando un paquete genético para que presente un dominio estructurado estable (el "*dominio de unión potencial inicial*", IPBD) para el que puede obtenerse una molécula de afinidad (que puede ser un anticuerpo). El éxito de las modificaciones se mide fácilmente determinando por ejemplo si el paquete genético modificado se une a la molécula de afinidad.

El IPBD se selecciona en vista de su tolerancia por la mutagénesis extensa. Una vez que se sabe que el IPBD puede presentarse en la superficie de un paquete y someterse a selección de afinidad, el gen que codifica el IPBD se somete a un patrón especial de múltiples mutagénesis, denominado en este documento "*variegación*", que después de unas etapas de clonación y amplificación apropiadas conduce a la producción de una población de paquetes genéticos cada uno de los cuales presenta un único dominio de unión potencial (un mutante del IPBD), pero que colectivamente presentan una multitud de dominios de unión potenciales diferentes aunque estructuralmente relacionados (PBD). Cada paquete genético tiene la versión del gen pbd que codifica el PBD que se presenta en la superficie de este paquete particular. Después se usa la selección por afinidad para identificar los paquetes genéticos que tienen los PBD con las características de unión deseadas y estos paquetes genéticos pueden amplificarse a continuación. Después de uno o más ciclos de enriquecimiento mediante selección por afinidad y amplificación, el ADN que codifica los dominios de unión exitosos (SBD) pueden recuperarse de los paquetes seleccionados.

Si fuera necesario, puede "variegarse" adicionalmente el ADN de los paquetes que tienen el SBD, usando un SBD del último ciclo de variegación como el "dominio de unión potencial parental" (PPBD) para la próxima generación de PBD, y el proceso puede continuar hasta que el especialista en la técnica esté satisfecho con el resultado. Por causa de las relaciones estructurales y evolutivas entre el IPBD y la primera generación de PBD, también se considera al IPBD como un "dominio de unión potencial parental" (PPBD).

Cuando se variegan micro-proteínas, los restos que están reticulados covalentemente en la molécula parental permanecen sin cambios, estabilizando de este modo la conformación. Por ejemplo, en la variegación de una micro-proteína unida con puentes disulfuro, determinadas cisteínas son invariables de modo que en las condiciones de expresión y presentación, se forman reticulaciones covalentes (por ejemplo, puentes disulfuro entre uno o más pares de cisteínas) y limitan sustancialmente la conformación que pueden adoptar los aminoácidos intermedios lineales hipervariables. En otras palabras, se introduce por manipulación genética un armazón limitante en los polipéptidos que de otro modo serían ampliamente aleatorizados.

Una vez que se ha caracterizado una micro-proteína de las características de unión deseadas, ésta puede producirse, no solamente por técnicas de ADN recombinante, sino también por métodos de síntesis no biológica.

Para los fines de las reivindicaciones adjuntas, una proteína P es una "proteína de unión" si para al menos una especie molecular, iónica o atómica A, diferente del dominio variable de un anticuerpo, la constante de disociación K_D (P,A) < 10^{-6} moles/litro (preferiblemente, < 10^{-7} moles/litro).

La exclusión de un "dominio variable de un anticuerpo" en (1) anteriormente pretende aclarar que para los fines de este documento no debe considerarse una proteína como "proteína de unión" simplemente porque sea antigénica.

La mayoría de las proteínas más grandes se pliegan en glóbulos distinguibles denominados dominios (ROSS81) Los dominios proteicos se han definido de diversas maneras; se prefieren las definiciones de "dominio" que hacen hincapié en la estabilidad - retención de la estructura global al afrontar fuerzas de perturbación tales como temperaturas elevadas o agentes caotrópicos -, aunque no se ignoran completamente los coordinados atómicos y la homología de la secuencia proteica.

Cuando un dominio de una proteína es principalmente responsable de la capacidad de la proteína para unirse específicamente a una diana seleccionada, se denomina en este documento "dominio de unión" (BD).

La expresión "ADN variegado" (ADNvg) se refiere a una mezcla de moléculas de ADN de la misma o de similar longitud que, cuando se alinean, varían en algunos codones para codificar en cada uno de dichos codones una pluralidad de diferentes aminoácidos, pero que solamente codifican un único aminoácido en el resto de posiciones del codón. Se entiende además que en el ADN variegado, los codones que son variables y el tamaño y la frecuencia de aparición de los diferentes aminoácidos que codifican un codón variable dado, se determinan por adelantado por el sintetizador del ADN, incluso aunque el método de síntesis no permita conocer, *a priori*, la secuencia de ninguna molécula de ADN individual en la mezcla. La cantidad de codones variables designados en el ADN variegado preferiblemente no es mayor de 20 codones, y más preferiblemente no es mayor de 5-10 codones. La mezcla de aminoácidos codificada en cada codón variable puede ser diferente de codón a codón.

Una población de paquetes genéticos en la que se ha introducido ADN variegado se dice que también está "variegada".

Para los fines de esta invención, la expresión "proteína de unión potencial" (PBP) se refiere a una proteína codificada por una especie de molécula de ADN en una población de ADN variegado en la que la región de variación aparece en una o más subsecuencias que codifican uno o más segmentos del polipéptido que tiene el potencial para servir como dominio de unión para la sustancia diana.

Una "proteína quimérica" es una fusión de una primera secuencia de aminoácidos (proteína) con una segunda secuencia de aminoácidos que define un domino extraño a y no sustancialmente homólogo con cualquier domino de la primera proteína. Una proteína quimérica puede presentar un dominio extraño que se haya descubierto (aunque en una proteína diferente) en un organismo que exprese también la primera proteína o puede ser una fusión "interespecífica" "intergenérica", etc. de estructuras proteicas expresadas por diferentes tipos de organismos.

Una secuencia de aminoácidos de las proteínas quiméricas de la presente invención se obtiene típicamente a partir de una proteína de la superficie externa de un "paquete genético" (GP) como se define en lo sucesivo en este documento. Uno que presenta un PBD en su superficie es un GP (PBD). La segunda secuencia de aminoácidos es una que, si se expresa en solitario, tendría las características de una proteína (o un dominio de la misma) pero se incorpora en la proteína quimérica como un dominio reconocible de la misma. Puede aparecer en el extremo amino o carboxilo de la primera secuencia de aminoácidos (sin ningún espaciador entre ambos), o puede interrumpir la primera secuencia de aminoácidos. La primera secuencia de aminoácidos puede corresponder exactamente con una proteína superficial del paquete genético, o puede modificarse por ejemplo para facilitar la presentación del dominio de unión.

II. *Micro- y otras mini-proteínas*

En la presente invención, se usan micro-proteínas unidas por puentes disulfuro como IPBD para verificar una estrategia de presentación y como PPBD para buscar la obtención de un BD con las características de unión a la diana deseadas. A menos que se indique otra cosa o lo requiera el contexto, las referencias en este documento a IPBD deben aplicarse, *mutatis mutandis*, también a PPBD.

Para los fines de las reivindicaciones adjuntas, una micro-proteína tiene entre aproximadamente seis y aproximadamente cuarenta restos; las micro-proteínas son un subconjunto de mini-proteínas, que tienen menos de aproximadamente sesenta restos. Puesto que las micro-proteínas forman un subconjunto de mini-proteínas, por motivos de comodidad, el término mini-proteínas se usará en ocasiones para referirse a micro-proteínas de la presente invención.

El IPBD puede ser una mini-proteína, con una actividad de unión conocida, o una que, aunque no posee una actividad de unión conocida, posee una estructura secundaria o superior que le proporciona por sí misma actividad de unión (hendiduras, surcos, etc.). Cuando el IPBD tiene una actividad de unión conocida, no es necesario que tenga ninguna actividad específica por el material diana. No es necesario que el IPBD sea idéntico en secuencia a una

mini-proteína de origen natural; puede ser un "homólogo" de una secuencia de aminoácidos que "se corresponde sustancialmente" con la de una mini-proteína conocida, o puede ser totalmente artificial.

Para determinar si debe considerarse que las secuencias "se corresponden sustancialmente", deben considerarse las siguientes cuestiones: el grado de similitud de la secuencia cuando las secuencias se alinean para un mejor ajuste de acuerdo con algoritmos convencionales, la similitud en los patrones de conectividad de cualquier reticulación (por ejemplo, puentes disulfuro), el grado en el que las proteínas tienen estructuras tridimensionales similares, como se indica mediante, por ejemplo, análisis de difracción de rayos X o RMN y el grado en el que las proteínas secuenciadas tienen actividad biológica similar. En este contexto, debe observarse que entre los inhibidores de serina proteasa existen familias de proteínas que se reconocen como homólogas puesto que hay pares de miembros con al menos el 30% de homología de la secuencia.

Un IPBD candidato debe cumplir los siguientes criterios:

1) que exista un dominio que permanecerá estable en las condiciones de su uso pretendido (el dominio puede comprender toda la proteína que se insertará, por ejemplo α-conotoxina GI (OLIV90a) o CMTI-III (MCWH89),

2) que pueda obtenerse el conocimiento de la secuencia de aminoácidos, y

3) que pueda obtenerse una molécula que tenga afinidad específica y alta por el IPBD, de forma abreviada AfM (IPBD).

Si solamente está disponible una especie de molécula que tiene afinidad por IPBD (AfM(IPBD)) se usará para: a) detectar el IPBD en la superficie del GP, b) optimizar el nivel de expresión y la densidad de la molécula de afinidad en la matriz, y c) determinar la eficacia y sensibilidad de la separación por afinidad. Sería preferible tener disponibles dos especies de AfM (IPBD), una con alta afinidad y otra con afinidad moderada por el IPBD. La especie con alta afinidad se usaría en la detección inicial y para determinar la eficacia y la sensibilidad, y la especie con afinidad moderada se usaría en la optimización.

Si el propio IPBD no es una proteína de unión conocida, o si su diana nativa no se ha purificado, puede usarse un anticuerpo generado contra el IPBD como la molécula de afinidad. El uso de un anticuerpo para este fin no debe interpretarse como que el anticuerpo es la diana definitiva.

Existen muchos IPBD candidatos para los cuales toda la información anterior está disponible o es razonablemente fácil de obtener, por ejemplo, CMTI-III (29 restos) (los inhibidores de tipo CMTI se describen en OTLE87, PAVE89, WIEC85, MCWH89, BODE89, HOLA89a,b), enterotoxina estable al calor (ST-Ia de *E. coli*) (18 restos) (GUAR89, BHAT86, SEKI85, SHIM87, TAKA85, TAKE90, THOM85a,b, YOSH85, DALL90, DWAR89, GARI87, GUZM89, GUZM90, HOUG84, KUBO89, KUPE90, OKAM87, OKAM88 y OKAM90), α-Conotoxina GI (13 restos) (HASH85, ALMQ89), μ-Conotoxina GIII (22 restos) (HID090) y la micro-proteína Conus King Kong (27 restos) (WOOD90). La información estructural puede obtenerse a partir de estudios de difracción de rayos X con neutrones, RMN, reticulación o marcado químico, modelado a partir de estructuras conocidas de proteínas relacionadas o a partir de cálculos teóricos. Se prefiere la información estructural 3D obtenida mediante difracción de rayos X, difracción de neutrones o RMN ya que estos métodos permiten la localización de casi todos los átomos dentro de límites definidos. La Tabla 50 muestra varios IPBD preferidos.

Las mutaciones pueden reducir la estabilidad del PBD. De este modo, el IPBD seleccionado debe tener preferiblemente una temperatura de fusión alta, por ejemplo de al menos 50ºC, y ser preferiblemente estable en un amplio intervalo de pH, por ejemplo de 8,0 a 3,0, pero más preferiblemente de 11,0 a 2,0, de modo que el SBD obtenido del IPBD seleccionado por mutación y selección a través de la unión, conservará la estabilidad suficiente. Preferiblemente, las sustituciones en el IPBD que producen los diversos PBD no reducen el punto de fusión del dominio por debajo de ~40ºC. Se entenderá que las mini-proteínas contienen reticulaciones covalentes, tales como uno o más puentes disulfuro y por lo tanto es probable que sean lo suficientemente estables.

In vitro, los puentes disulfuro pueden formarse espontáneamente en polipéptidos como resultado de la oxidación al aire. Este asunto es más complicado *in vivo*. Muy pocas proteínas intracelulares tienen puentes disulfuro, probablemente debido a que se mantiene un entorno fuertemente reductor por el sistema deL glutatión. Los puentes disulfuro son comunes en proteínas que viajan u operan en espacios intracelulares, tales como venenos de serpiente y otras toxinas (por ejemplo, conotoxinas, caribdotoxinas, enterotoxinas bacterianas), hormonas peptídicas, enzimas digestivas, proteínas complemento, inmunoglobulinas, lisozimas, inhibidores de proteasa (BPTI y sus homólogos, CMTI-III (inhibidor de la tripsina III de *Cucurbita maxima*) y sus homólogos, hirudina, *etc*.) y proteínas de la leche.

Se han descubierto puentes disulfuro que cierran estrechamente los bucles intracatenarios en pepsina, tiorredoxina, en la cadena A de la insulina, fibroína de la seda y lipoamida deshidrogenasa. Los restos de cisteína que forman puentes están separados por uno a cuatro restos a lo largo de la cadena polipeptídica. La construcción de modelos, análisis de difracción de rayos X y estudios de RMN han demostrado que la estructura de carbono de tales bucles es normalmente plana y rígida.

Existen dos tipos de puentes disulfuro en las inmunoglobulinas. Uno es el puente intracatenario conservado, que se extiende de aproximadamente 60 a 70 restos aminoacídicos y que se encuentra, repetidamente, en casi todos los dominios de inmunoglobulina. Enterrados profundamente entre las láminas β opuestas, estos puentes están protegidos del disolvente y normalmente pueden reducirse solamente en presencia de agentes desnaturalizantes. El resto de puentes disulfuro son principalmente puentes intercatenarios que se sitúan en la superficie de la molécula; son accesibles al disolvente y se reducen de forma relativamente fácil (STEI85). Los puentes disulfuro de las micro-proteínas de la presente invención son enlaces intracatenarios entre cisteínas que tienen espaciadores en la cadena mucho más pequeños.

Las micro-proteínas de la presente invención contienen una pluralidad de enlaces disulfuro; es preferible que al menos dos cisteínas estén agrupadas, es decir, que estén inmediatamente adyacentes a lo largo de la cadena (-C-C-) o estén separadas por un solo aminoácido (-C-X-C-). En ambos casos, las dos cisteínas agrupadas se vuelven incapaces de emparejarse entre sí por razones estéricas y se reduce el número de topologías factibles.

Un puente disulfuro intracatenario que conecta los aminoácidos 3 y 8 de un polipéptido de 16 restos se designará en este documento como que tiene un tramo de 4. Si los aminoácidos 4 y 12 también están unidos por un puente disulfuro, entonces forman un segundo tramo de 7. En conjunto, las cuatro cisteínas dividen el péptido en cuatro segmentos intercisteína (1-2, 5-7, 9-11 y 13-16). (Véase que no existe segmento entre Cys3 y Cys4). El patrón de conectividad de una micro-proteína reticulada es una sencilla descripción de la situación relativa de los extremos de las reticulaciones. Por ejemplo, para una micro-proteína con dos puentes disulfuro, el patrón de conectividad "1-3, 2-4" significa que la primera cisteína reticulada está unida por un puente disulfuro a la tercera cisteína reticulada (en la secuencia primaria), y la segunda a la cuarta.

El grado en el que la reticulación limita la libertad conformacional de la mini-proteína y el grado en el que ésta estabiliza la mini-proteína puede evaluarse mediante varios medios. Éstos incluyen espectroscopía de absorción (que puede mostrar si un aminoácido está oculto o expuesto), estudios de dicroísmo circular (que proporcionan una representación general del contenido helicoidal de la proteína), formación de imágenes por resonancia magnética nuclear (que muestra la cantidad de núcleos en un entorno químico particular así como la movilidad de los núcleos) y análisis de difracción de rayos X o de neutrones de cristales de proteínas. La estabilidad de la mini-proteína puede determinarse controlando los cambios en la absorción a diversas longitudes de onda en función de la temperatura, pH, *etc*.; los restos ocultos quedan expuestos a medida que la proteína se despliega. Análogamente, el desplegamiento de la proteína como resultado de condiciones desnaturalizantes provoca cambios en las posiciones lineales y las anchuras de RMN. Los espectros de dicroísmo circular (CD) son extremadamente sensibles a la conformación.

Las micro-proteínas unidas por puentes disulfuro variegadas de la presente invención son:

Micro-proteínas de Clase I que presentan un único par de cisteínas capaces de interaccionar para formar un puente disulfuro, teniendo dicho puente un tramo de no más de aproximadamente nueve restos. Este puente disulfuro tiene un tramo de al menos dos restos; esto está en función de la geometría del puente disulfuro. Cuando la separación es de dos o tres restos, un resto puede ser glicina para reducir la tensión de los restos unidos por el puente. El límite superior de separación es menos preciso, sin embargo en general, cuando mayor sea la separación, menor será la limitación en la conformación impuesta sobre los restos de aminoácidos lineales intermedios por el puente disulfuro.

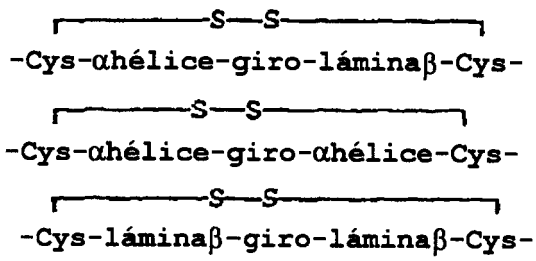
La cadena principal de dicho péptido tiene muy poca libertad, pero no se somete a tensión. La energía libre liberada cuando se forma el puente disulfuro supera la energía libre perdida por la cadena principal cuando se cierra en una conformación que junta las cisteínas. Al haber perdido la energía libre de la formación del puente disulfuro, los extremos proximales de los grupos laterales se mantienen en una relación más o menos fija entre sí. Cuando se une a una diana, el dominio no necesita gastar energía libre para alcanzar la conformación correcta. El dominio no puede saltar a otra conformación y unirse a una no diana.

Un puente disulfuro con un tramo de 4 ó 5 es especial. Si el tramo se aumenta a 6, la influencia de limitación se reduce. El menos uno de los aminoácidos incluido puede ser un aminoácido que imponga restricciones sobre la geometría de la cadena principal. La mayor restricción la impone la prolina. La valina y la isoleucina restringen la cadena principal en menor medida. La posición preferida para este resto no cisteína limitante es adyacente a una de las cisteínas invariables, sin embargo, puede estar en cualquiera de los otros restos unidos por el puente. Si el tramo es de siete, pueden incluirse dos aminoácidos que limiten la conformación de la cadena principal. Estos aminoácidos pueden estar en cualquiera de las siete posiciones, pero pueden ser los dos restos unidos por el puente que están inmediatamente adyacentes a las cisteínas. Si el tramo es de ocho o nueve, pueden proporcionarse aminoácidos limitantes adicionales.

Aunque una micro-proteína de clase I puede tener hasta 40 aminoácidos, más preferiblemente no tiene más de 20 aminoácidos.

El puente disulfuro de una micro-proteína de clase I está expuesto al disolvente. De este modo, debe evitarse normalmente exponer la población variegada de GP que presentan micro-proteínas de Clase I a reactivos que rompen puentes disulfuro.

Las *micro-proteínas de Clase II* son aquellas que presentan un único puente disulfuro que tiene un tramo de más de nueve aminoácidos. Los aminoácidos unidos por puente forman estructuras secundarias que ayudan a estabilizar su conformación. Estos aminoácidos intermedios pueden formar estructuras súper secundarias en forma de horquilla tales como las que se esquematizan a continuación:



En base a estudios de proteínas conocidas, puede calcularse la propensión de un resto en particular o de un dipéptido o tripéptido particular para encontrarse en una α hélice, lámina β o giro inverso. Las frecuencias normalizadas de aparición de los restos de aminoácidos en estas estructuras secundarias se proporcionan en la Tabla 6-4 de CRBI84. Para un tratamiento más detallado sobre la predicción de la estructura secundaria a partir de la secuencia de aminoácidos, véase el Capítulo 6 de SCHU79.

Al diseñar una estructura de horquilla adecuada, puede copiarse una estructura real de una proteína cuya conformación tridimensional se conozca, diseñar la estructura usando datos de frecuencia o combinar los dos enfoques. Pueden usarse una o más estructuras reales como modelo y los datos de frecuencia pueden usarse para determinar qué mutaciones pueden realizarse sin alterar la estructura.

No puede haber más de tres aminoácidos entre la cisteína y el comienzo o el final de la α hélice o lámina β.

También son posibles estructuras más complejas (tales como doble horquilla).

Las *micro-proteínas de Clase IIIa* son aquellas que presentan dos puentes disulfuro. También pueden presentar opcionalmente estructuras secundarias tales como las que se han analizado anteriormente con respecto a las micro-proteínas de Clase II. Con dos puentes disulfuro existen tres topologías posibles; si se desea, la cantidad de topologías de puentes disulfuro realizables puede reducirse agrupando las cisteínas como en la enterotoxina ST-Ia termoestable.

Las *micro-proteínas de Clase IIIb* son aquellas que presentan tres o más puentes disulfuro y preferiblemente al menos una agrupación de cisteínas como se ha descrito anteriormente.

Mini-proteínas de Dedo Metálico. Las mini-proteínas pueden no reticularse mediante enlaces disulfuro, por ejemplo, análogos de proteínas tipo dedo. Las proteínas tipo dedo se caracterizan por estructuras de tipo dedo en las que se coordina un ión metálico mediante dos restos Cys y dos restos His, formando una disposición tetraédrica a su alrededor. El ión metálico más frecuente es cinc (II), pero puede ser hierro, cobre, cobalto, *etc*. El "dedo" tiene la secuencia consenso (Phe o Tyr)-(1 AA)-Cys-(2-4 AAs)-Cys-(3 AAs)-Phe-(5 AAs)-Leu-(2 AAs)-His-(3 AA)-His-(5 AAs) (HERG88; GIBSB8). Aunque las proteínas tipo dedo contienen típicamente muchas repeticiones del motivo dedo, se sabe que un único dedo se plegará en presencia de iones cinc (FRAN87; PARR88). Existe una controversia acerca de si son necesarios dos dedos para unirse al ADN. La expresión Mini-proteínas de Dedo Metálico abarca mini-proteínas con uno o dos dedos. Otras combinaciones de grupos laterales pueden conducir a la formación de reticulaciones que implican iones metálicos multivalentes. Summers (SUMM91), por ejemplo, presenta una mini-proteína de 18 aminoácidos descubierta en la proteína de la cápsida del HIV-1-F1 y que tiene tres cisteínas y una histidina que se une a un átomo de cinc. Debe entenderse que no es necesario que la diana sea un ácido nucleico.

G. *PBS modificados*

Existen varias enzimas y reactivos químicos que pueden modificar de forma selectiva algunos grupos laterales de las proteínas, incluyendo: a) protein-tirosina quinasa, reactivo de Ellman, metiltransferasas (que metilan grupos laterales GLU), serina quinasas, prolina hidroxilasas, enzimas dependientes de la vitamina K que convierten GLU en GLA, anhídrido maleico y reactivos alquilantes. El tratamiento de la población variegada de GP(PBD) con una de estas enzimas o reactivos modificará los grupos laterales afectados por la enzima o reactivo seleccionado. Se prefieren las enzimas y reactivos que no destruyen el GP. Dichas modificaciones de grupos laterales pueden afectar directamente a las propiedades de unión de los PBD presentados. Usando métodos de separación por afinidad, se enriquecen los GP modificados que se unen a la diana predeterminada. Puesto que el dominio de unión activo no está totalmente especificado genéticamente, debe repetirse la modificación post-morfogénesis en cada ronda de enriquecimiento. Este enfoque es particularmente apropiado con IPBD de mini-proteínas porque se prevé la síntesis química de estos SBD.

III. *Estrategia de variegación - mutagénesis para obtener dominios de unión potenciales con la diversidad deseada*

III.A. *En Líneas Generales*

Cuando la cantidad de secuencias de aminoácidos diferentes que pueden obtenerse mediante la mutación del dominio es grande en comparación con la cantidad de dominios diferentes que pueden presentarse en cantidades detectables, la eficacia de la evolución forzada se potencia enormemente mediante la elección cuidadosa de los restos que se modificarán. En primer lugar, se identifican los restos de una proteína conocida que es probable que afecten a su actividad de unión (por ejemplo, restos de superficie) y no es probable que degraden indebidamente su estabilidad. Después, se modifican simultáneamente todos o parte de los codones que codifican estos restos para producir una población de ADN variegada. Los grupos de restos superficiales que están lo suficientemente cerca entre sí en la superficie para entrar en contacto con una molécula diana simultáneamente son conjuntos preferidos para la variegación simultánea. La población de ADN variegada se usa para expresar diversos dominios de unión potenciales, cuya capacidad para unirse a la diana de interés puede evaluarse después.

De este modo, el proceso de la presente invención se distingue adicionalmente de otros métodos por la naturaleza de la población altamente variegada que se produce y de la que se seleccionan nuevas proteínas de unión. Se fuerza al dominio de unión potencial presentado a muestrear la "secuencia espaciadora" cercana de las secuencias de aminoácidos relacionadas de manera eficaz y organizada. Cuatro objetivos guían los diversos planes de variegación usados en este documento, preferiblemente: 1) está disponible una gran cantidad (por ejemplo, 10^7) de variantes, 2) aparece realmente un porcentaje muy alto de las variantes posibles en cantidades detectables, 3) la frecuencia de aparición de las variantes deseadas es relativamente uniforme y 4) la variación se produce solamente en una cantidad limitada de restos de aminoácidos, más preferiblemente en restos que tienen grupos laterales dirigidos hacia una región común en la superficie del dominio de unión potencial.

Debe distinguirse del uso aislado de agentes mutagénicos indiscriminados, tales como radiación e hidroxilamina para modificar un gen, donde no hay control (o éste es de forma muy indirecta) sobre el sitio de mutación. Muchas de las mutaciones afectarán a restos que no forman parte del dominio de unión. Cuando los mutágenos químicos se dirigen hacia todo el genoma, la mayoría de las mutaciones se producen en genes diferentes al que codifica el dominio de unión potencial. Además, puesto que a un nivel de mutagénesis razonable, cualquier codón modificado es probable que se caracterice por un único cambio de bases, solamente se explorarán un intervalo limitado y poco imparcial de posibilidades. Igualmente distante está el uso de técnicas de mutagénesis específicas para un sitio que emplean oligonucleótidos mutagénicos de secuencia no aleatorizada, ya que estas técnicas no proporcionan por sí mismas la producción y el ensayo de gran cantidad de variantes. Aunque se conocen técnicas centradas en la mutagénesis aleatoria, la importancia del control de la distribución de la variación se ha pasado por alto ampliamente.

Los dominios de unión potenciales se han diseñado en primer lugar a nivel de los aminoácidos. Una vez que se ha identificado qué restos deben mutagenizarse y qué mutaciones se permitirán en estas posiciones, puede diseñarse después el ADN variegado que codificará los diversos PBD para asegurar que haya una probabilidad razonable de que si el PBD tiene afinidad por la diana, éste se detecte. Por supuesto, la cantidad de transformantes independientes obtenidos y la sensibilidad de las tecnologías de separación por afinidad impondrán límites al grado de variegación posible dentro de cualquier turno sencillo de variegación.

Existen muchas maneras de generar diversidad en una proteína (véase RICH86, CARU85 y OLIP86). Por un lado, se modifican unos pocos restos de la proteína en la medida de lo posible (*inter alia*, véase CARU85, CARU87, RICH86 y WHAR86). Este enfoque se denominará "Mutagénesis Centrada". Una estrategia de "Mutagénesis Centrada" típica es tomar un conjunto de cinco a 7 restos y modificar cada uno de 13 a 20 posibilidades. Un plan alternativo de mutagénesis ("Mutagénesis Difusa") es modificar más restos a través de un conjunto de elecciones más limitado (véase VERS86a y PAKU86). El patrón de variegación adoptado puede estar entre estos extremos, *por ejemplo*, dos restos modificados entre los veinte aminoácidos, dos más entre solamente dos posibilidades y un quinto entre diez de los veinte aminoácidos.

No existe un límite fijado de la cantidad de codones que pueden mutarse simultáneamente. Sin embargo, es deseable adoptar una estrategia de mutagénesis que dé como resultado una probabilidad razonable de que, de hecho, se presente una secuencia de PBD posible en al menos un paquete genético. Preferiblemente, la probabilidad de que una muteína codificada por el ADNvg y compuesta por los aminoácidos menos favorecidos en cada posición variegada se presente en al menos un transformante independiente de la biblioteca es de al menos 0,50 y más preferiblemente de al menos 0,90. (Por supuesto, es más probable que las muteínas compuestas por aminoácidos más favorecidos se produzcan en la misma biblioteca).

Preferiblemente, la variegación de este modo provocará que una población transformante típica presente 10^6-10^7 secuencias de aminoácidos diferentes por medio de preferiblemente no más de 10 veces más (más preferiblemente, no más de 3 veces) secuencias de ADN diferentes.

Para las micro-proteínas de Clase III, se variegarán preferiblemente no más de 20 y más preferiblemente 5-10 codones. Sin embargo, si se emplea la mutágenesis difusa, el número de codones a variegar puede ser superior.

Aunque la variegación implicará normalmente la sustitución de un aminoácido por otro en un codón designado variable, también puede implicar la inserción o deleción de aminoácidos.

III.B. *Identificación del Resto a Modificar*

A continuación se consideran los principios que guían la elección de restos del IPBD a modificar. Un concepto clave es que solamente las proteínas estructuradas muestran unión específica, *es decir* pueden unirse a una entidad química particular para excluir a la mayoría de las demás. De este modo, los restos a modificar se seleccionan teniendo en cuenta la preservación de la estructura del IPBD subyacente. Las sustituciones que previenen el plegamiento del PBD provocarán GP que tienen aquellos genes que se unen indiscriminadamente de modo que puedan retirarse fácilmente de la población. Es menos probable que las sustituciones de aminoácidos que están expuestos al disolvente afecten a la estructura 3D que las sustituciones en loci internos. (Véase PAKU86, REID88a, EISE85, SCHU79, págs. 169-171 y CREI84, págs. 239-245, 314-315). Los restos internos se conservan frecuentemente y el tipo de aminoácido no puede cambiar a un tipo significativamente diferente sin un riesgo sustancial de que la estructura de la proteína se altere. Sin embargo, se toleran algunos cambios conservativos de restos internos, tales como I a L o F a Y. Dichos cambios conservativos afectan ligeramente a la colocación y a la dinámica de los restos de la proteína adyacente y tal "ajuste preciso" puede ser útil una vez que se ha descubierto un SBD. Las inserciones y las deleciones se toleran más fácilmente bucles que en cualquier otro lugar. (THOR88).

Los datos acerca del IPBD y la diana que son útiles para decidir qué residuos se modificarán en el ciclo de variegación incluyen: 1) estructura 3D o al menos una lista de restos en la superficie del IPBD, 2) lista de secuencias homólogas al IPBD y 3) modelo de la molécula diana o un sustituto de la diana.

III.C. *Determinación del Conjunto de Sustitución para cada Resto Parental*

Habiendo tomado los restos a modificar, se decide ahora el intervalo de aminoácidos permitido para cada resto variable. El nivel total de variegación es el producto de la cantidad de variantes en cada resto modificado. Cada resto modificado puede tener un esquema diferente de variegación, produciendo de 2 a 20 posibilidades diferentes. El conjunto de aminoácidos que están codificados potencialmente por un codón variegado dado se denomina su "conjunto de sustitución".

El ordenador que controla un sintetizador de ADN, tal como el Milligen 7500, puede programarse para sintetizar cualquier base de un oligonucleótido con cualquier distribución de nucleótidos tomando algunos sustratos de nucleótidos (*por ejemplo*, nucleótidos fosforamidados) de cada una de dos o más reservas. Como alternativa, pueden mezclarse sustratos de nucleótidos en cualquier proporción y colocarse en una de las reservas extraordinarias para la llamada síntesis "de frasco sucio". Cada codón podría programarse de forma diferente. La "mezcla" de bases en cada posición del nucleótido del codón determina la frecuencia relativa de aparición de los diferentes aminoácidos codificados por este codón.

Los codones variegados de forma simple son aquellos en los que las posiciones del nucleótido que degeneran se obtienen a partir de una mezcla de dos o más bases mezcladas en proporciones equimolares. Estas mezclas se describen en esta memoria descriptiva por medio del código convencional de "nucleótido ambiguo". En este código, por ejemplo, en el codón degenerado "SNT", "S" significa una mezcla equimolar de bases G y C, "N", una mezcla equimolar de las cuatro bases, y "T", la única e invariable base timidina.

Los codones variegados de forma compleja son aquellos en los que al menos una de las tres posiciones está ocupada por una base diferente de una mezcla equimolar de dos o más bases.

Pueden usarse codones variegados de forma simple o compleja para conseguir el conjunto de sustitución deseado.

Si no se tiene información que indique que un aminoácido o clase de aminoácidos particular es apropiada, se intentan sustituir todos los aminoácidos con la misma probabilidad, ya que la presentación de una mini-proteína por encima del nivel detectable es inútil. Las cantidades iguales de los cuatro nucleótidos en cada posición en un codón (NNN) produce la distribución de aminoácidos en la que cada aminoácido está presente en una proporción con respecto al número de codones que lo codifican. Esta distribución tiene la desventaja de proporcionar dos restos básicos por cada resto ácido. Además, se producen hasta seis veces más R, S y L que W o M. Si se sintetizan cinco codones con esta distribución, cada una de las 243 secuencias que codifican alguna combinación de L, R y S son 7776 veces más abundantes que cada una de las 32 secuencias que codifican alguna combinación de W y M. Para tener cinco W presentes a niveles detectables, debemos tener cada una de las secuencias (L, R, S) en un exceso de 7776 veces.

Los restos de aminoácidos particulares pueden influir en la estructura terciaria de un polipéptido definido de varias maneras, incluyendo:

a) afectando la flexibilidad de la cadena principal del polipéptido.

b) añadiendo grupos hidrófobos,

c) añadiendo grupos cargados,

d) permitiendo la formación de puentes de hidrógeno, y

e) formando reticulaciones, tales como puentes disulfuro, quelando iones metálicos o uniendo grupos prostéticos.

Lundeen (LUND86) ha presentado en una tabla las frecuencias de aminoácidos en hélices, cadenas β, giros y enrollamientos en proteínas de estructura 3D conocida y ha distinguido entre CYS que tienen grupos tioles libres y semicisteínas. Informa que la CYS libre se encuentra más a menudo en hélices mientras que las semicisteínas se encuentran más a menudo en láminas β. Sin embargo, las semicisteínas se encuentran regularmente en las hélices. Pease *et al*. (PEAS90) construyeron un péptido que tenía dos cisteínas; un extremo de cada una de las cuales es una α hélice muy estable. La apamina tiene una estructura similar (WEMM83, PEAS88).

Flexibilidad:

GLY es el aminoácido más pequeño que tiene dos hidrógenos unidos al C_α. Puesto que GLY no tiene C_β, proporciona la mayor flexibilidad a la cadena principal. De este modo, GLY aparece muy frecuentemente en giros inversos, particularmente junto con PRO, ASP, ASN, SER y THR.

Los aminoácidos ALA, SER, CYS, ASP, ASN, LEU, MET, PHE, TYR, TRP, ARG, HIS, GLU, GLN y LYS tienen carbonos β sin ramificar. De éstos, los grupos laterales de SER, ASP y ASN forman frecuentemente puentes de hidrógeno con la cadena principal y pueden tomar conformaciones de la cadena principal que son desfavorables energéticamente para el resto. VAL, ILE y THR tienen carbonos ramificados que hacen a la conformación de la cadena principal extendida más favorable. De este modo, VAL e ILE se observan más frecuentemente en láminas β. Puesto que el grupo lateral de THR puede formar puentes de hidrógeno fácilmente con la cadena principal, tiene menos tendencia a aparecer en una lámina β.

La cadena principal de prolina está limitada particularmente por el grupo lateral cíclico. El ángulo ϕ está siempre cercano a 60º. La mayoría de las prolinas se encuentran cerca de la superficie de la proteína.

Carga:

LYS y ARG tienen una única carga positiva a cualquier pH por debajo de 10,4 ó 12,0 respectivamente. Sin embargo, los grupos metileno, cuatro y tres respectivamente, de estos aminoácidos, son capaces de realizar interacciones hidrófobas. El grupo guanidinio de ARG es capaz de donar cinco hidrógenos simultáneamente, mientras que el grupo amino de LYS puede donar solamente tres. Además, la geometría de estos grupos es bastante diferente, de modo que estos grupos a menudo no son intercambiables.

ASP y GLU tienen una única carga negativa a cualquier pH por encima de ~4,5 y 4,6 respectivamente. Puesto que ASP tiene solamente un grupo metileno, son posibles unas pocas interacciones hidrófobas. La geometría de ASP le permite a él mismo formar puentes de hidrógeno con nitrógenos de la cadena principal, lo que es coherente con que el ASP se encuentre muy a menudo en los giros inversos y al comienzo de las hélices. GLU se encuentra más a menudo en α hélices y particularmente en la porción amino-terminal de estas hélices ya que la carga negativa del grupo lateral tiene una interacción estabilizante con el dipolo de la hélice (NICH88, SALI88).

HIS tiene un pK de ionización en el intervalo fisiológico, concretamente, 6,2. Este pK puede alterarse por la cercanía de grupos cargados o de donadores o aceptores de hidrógeno. HIS es capaz de formar puentes con iones metálicos tales como cinc, cobre y hierro.

Puentes de hidrógeno:

Además de los aminoácidos cargados, SER, THR, ASN, GLN, TYR y TRP pueden participar en los puentes de hidrógeno.

Reticulaciones:

La forma más importante de reticulación es el puente disulfuro formado entre los tioles de restos CYS. En un entorno oxidante adecuado, estos puentes se forman espontáneamente. Estos puentes pueden estabilizar enormemente una conformación particular de una proteína o mini-proteína. Cuando está presente una mezcla de reactivos de tiol oxidado y reducido, tienen lugar reacciones de intercambio que permiten que predomine la conformación más estable. Con respecto a los puentes disulfuro en proteínas y péptidos, véase también KATZ90, MATS89, PERR84, PERR86, SAUE86, WELL86, JANA89, HORV89, KISH85 y SCHN86.

Otras reticulaciones que se forman sin necesidad de enzimas específicas incluyen:

1)	$(CYS)_4$:Fe	Rubredoxina (en CREI84, P.376)
2)	$(CYS)_4$:Zn	Aspartato Transcarbamilasa (en CREI84, P.376) y dedos de Zn (HARD90)
3)	$(HIS)_2$(MET)(CYS):Cu	Azurina (en CREI84, P.376) y proteína de Pepino de Cobre "Azul" Básico (GUSS88)
4)	$(HIS)_4$:Cu	CuZn superóxido dismutasa
5)	$(CYS)_4$:(Fe_4S_4)	Ferredoxina (en CREI84, P.376)
6)	$(CYS)_2(HIS)_2$:Zn	Dedos de cinc (GIBS88, SUMM91)
7)	$(CYS)_3(HIS)$:Zn	Dedos de cinc (GAUS87, GIBS88)

Las reticulaciones que tienen $(HIS)_2$ (MET) (CYS):Cu tienen la ventaja potencial de que HIS y MET no pueden formar otras reticulaciones sin Cu.

Codones Variegados de Forma Sencilla

Los siguientes codones variegados de forma sencilla son útiles puesto que codifican un conjunto de aminoácidos relativamente equilibrado:

1) SNT que codifica el conjunto [L,P,H,R,V,A,D,G]: a) uno ácido (D) y uno básico (R), b) los dos alifáticos (L,V) y aromáticos hidrófobos (H), c) grupos laterales grandes (L,R,H) y pequeños (G,A), d) aminoácidos rígidos (P) y flexibles (G), e) cada aminoácido codificado una vez.

2) RNG que codifica el conjunto [M, T, R, R, V, A, E, G]: a) uno ácido y dos básicos (no óptimo, pero aceptable), b) hidrófilos e hidrófobos, c) cada aminoácido codificado una vez.

3) RMG que codifica el conjunto [T,K,A,E]: a) uno ácido, uno básico, uno neutro hidrófilo, b) tres α hélices favorables, c) cada aminoácido codificado una vez.

4) VNT que codifica el conjunto [L,P,H,R,I,T,N,S,V,A,D,G]: a) uno ácido, uno básico, b) todas las clases: cargados, neutros, hidrófilos, hidrófobos, rígidos y flexibles, etc., c) cada aminoácido codificado una vez.

5) RRS que codifica el conjunto [N,S,K,R,D,E,G^2]: a) dos ácidos, dos básicos, b) dos neutros hidrófilos, c) solamente la glicina codificada dos veces.

6) NNT que codifica el conjunto [F,S,Y,C,L,P,H,R,I,T,N,V,A-,D,G]: a) dieciséis secuencias de ADN proporcionan quince aminoácidos diferentes; solamente se repite la serina, todos los demás están presentes en cantidades iguales (esto permite un muestreo muy eficaz de la biblioteca), b) hay cantidades iguales de aminoácidos ácidos y básicos (D y R, uno cada vez), c) todas las clases principales de aminoácidos están presentes: ácidos, básicos, alifáticos hidrófobos, aromáticos hidrófobos y neutros hidrófilos.

7) NNG, que codifica el conjunto [L^2,R^2,S,W,P,Q,M,T,R,V,A,-E,G, detención]; a) una preponderancia clara de restos que favorecen la formación de α-hélices [L,M,A,Q,K,E; y en menor medida, S,R,T]; b) codifica 13 aminoácidos diferentes. (VHG codifica un subconjunto del conjunto codificado por NNG que codifica 9 aminoácidos en nueve secuencias de ADN diferentes, con igual cantidad de ácidos y bases y 5/9 que favorecen la α-hélice).

Para la variegación inicial, en la mayoría de los casos se prefiere NNT. Sin embargo, cuando el codón codifica un aminoácido para incorporar en una α hélice, se prefiere NNG.

A continuación, se analizan varias variegaciones sencillas en referencia a la eficacia con la que pueden muestrearse las bibliotecas.

Se han presentado bibliotecas de hexapéptidos aleatorios codificados por $(NNK)^6$ (SCOT90, CWIR90). La Tabla 130 muestra el comportamiento esperado de dichas bibliotecas. NNK produce codones sencillos para PHE, TYR, CYS, TRP, HIS, GLN, ILE, MET, ASN, LYS, ASP y GLU (conjunto α); dos codones para cada uno de VAL, ALA,

PRO, THR y GLY (conjunto ϕ); y tres codones para cada uno de LEU, ARG y SER (conjunto Ω). Se han dividido las 64.000.000 de secuencias posibles en 28 clases, que se muestran en la Tabla 130A, en base a la cantidad de aminoácidos de cada uno de estos conjuntos. La clase más grande es $\phi\Omega\alpha\alpha\alpha$ con ~14,6% de las secuencias posibles. Aparte de cualquier selección, todas las secuencias en una clase tienen la misma probabilidad de producirse. La Tabla 130B muestra la probabilidad de que una secuencia de ADN dada tomada de la biblioteca $(NNK)^6$ codifique un hexapéptido que pertenezca a una de las clases definidas; obsérvese que solamente el ~6,3% de las secuencias de ADN pertenecen a la clase $\phi\Omega\alpha\alpha\alpha$.

La Tabla 130C muestra las cantidades esperadas de secuencias en cada clase para bibliotecas que contienen diversas cantidades de transformantes independientes (*concretamente*, 10^6, $3{\cdot}10^6$, 10^7, $3{\cdot}10^7$, 10^8, $3{\cdot}10^8$, 10^9 y $3{\cdot}10^9$). Con 10^6 transformantes independientes (IT), se espera observar el 56% de la clase $\Omega\Omega\Omega\Omega\Omega\Omega$, pero solamente el 0,1% de la clase $\alpha\alpha\alpha\alpha\alpha\alpha$. La amplia mayoría de secuencias observadas vienen de clases para las que se muestrea menos del 10% de la clase. Se supone que un péptido por ejemplo de la clase $\phi\phi\Omega\Omega\alpha\alpha$ se aísla fraccionando la biblioteca para la unión a una diana. Se considera cuánto se conoce acerca de los péptidos que están relacionados con la secuencia aislada. Puesto que solamente se muestreó el 4% de la clase $\phi\phi\Omega\Omega\alpha\alpha$, no puede concluirse que los aminoácidos del conjunto Ω sean de hecho los mejores del conjunto Ω. Podemos tener LEU en la posición 2, pero ARG o SER podrían ser mejores. Incluso si se aísla un péptido de la clase $\Omega\Omega\Omega\Omega\Omega\Omega$, hay una marcada posibilidad de que los mejores miembros de la clase no estén presentes en la biblioteca.

Con una biblioteca de 10^7 IT, se observa que se han muestreado completamente varias clases, pero la clase $\alpha\alpha\alpha\alpha\alpha\alpha$ se ha muestreado solamente en un 1,1%. A $7{,}6{\cdot}10^7$ IT, se espera la presentación del 50% de todas las secuencias de aminoácidos, pero las clases que contienen tres o más aminoácidos del conjunto siguen estando poco muestreadas. Para conseguir el muestreo completo de la biblioteca $(NNK)^6$ se necesitan aproximadamente $3{\cdot}10^9$ IT, 10 veces más grandes que la biblioteca $(NNK)^6$ más grande presentada hasta ahora.

La Tabla 131 muestra las expectativas para una biblioteca codificada por $(NNT)^4(NNG)^2$. Las expectativas de abundancia son independientes del orden de los codones o de los codones no modificados intercalados. Esta biblioteca codifica 0,133 veces como tantas secuencias de aminoácidos, pero solamente 0,0165 veces como tantas secuencias de ADN. Por lo tanto, $5{,}0{\cdot}10^7$ IT (es decir, 60 veces menos de las necesarias para $(NNK)^6$) proporcionan un muestreo casi completo de la biblioteca. Los resultados serían ligeramente mejores para $(NNT)^6$ y ligeramente, pero no mucho, peores para $(NNG)^6$. El factor de control es la proporción de secuencias de ADN con respecto a secuencias de aminoácidos.

La Tabla 132 muestra la proporción de Nº de secuencias de ADN/Nº de secuencias de aminoácidos para los codones NNK, NNT y NNG. Para NNK y NNG se supone que el PBD se presenta como parte de un gen esencial, tal como un gen *III* en el fago Ff, como se indica mediante la expresión "se supone que las detenciones desaparecen". No se necesita de ninguna manera que se use dicho gen esencial. Si se usa un gen no esencial, el análisis sería ligeramente diferente; el muestreo de NNK y NNG sería ligeramente menos eficaz. Obsérvese que $(NNT)^6$ proporciona 3,6 veces más secuencias de aminoácidos que $(NNK)^5$ pero requiere 1,7 veces *menos* secuencias de ADN. Véase también que $(NNT)^7$ da hasta *dos veces* más secuencias de aminoácidos que $(NNK)^6$, pero 3,3 veces *menos* secuencias de ADN.

De este modo, aunque es posible usar una mezcla sencilla (NNS, NNK o NNN) para obtener en una posición particular los veinte aminoácidos, estas mezclas sencillas conducen a un conjunto altamente imparcial de aminoácidos codificados. Este problema puede superarse mediante el uso de de codones variegados de forma compleja.

Codones variegados de forma compleja

La distribución de nucleótidos ("fxS") dentro del codón que permite los veinte aminoácidos y que produce la mayor proporción de abundancia de los aminoácidos menos favorecidos (Ifaa) con respecto a la de los aminoácidos más favorecidos (mfaa), sometido a las limitaciones de abundancias equivalentes de aminoácidos ácidos y básicos, la menor cantidad posible de codones de detención y, por motivos de comodidad, siendo la tercera base T o G, se muestra en la Tabla 10A y produce moléculas de ADN que codifican cada tipo de aminoácido con las abundancias que se muestran. Otros codones variegados de forma compleja se obtienen suavizando una o más limitaciones.

Véase que esta construcción química modifica los veinte aminoácidos, siendo los aminoácidos ácidos y básicos equiprobables, y el aminoácido más favorecido (serina) se codifica solamente 2,454 veces más a menudo que el aminoácido menos favorecido (triptófano). El codón "fxS" vg mejora más el muestreo para los péptidos que contienen varios de los aminoácidos [F,Y,C,W,H-,Q,I,M,N,K,D,E] para los que NNK o NNS proporcionan solamente un codón. Estas ventajas de muestreo son más pronunciadas cuando la biblioteca es relativamente pequeña.

Los resultados de la omisión de los requisitos de igualdad de ácidos y bases y de minimizar los codones de detención se muestran en la Tabla 10B.

Las ventajas de un codón NNT se analizan además en la presente solicitud. El NNT no optimizado proporciona 15 aminoácidos codificados solamente por 16 secuencias de ADN. Es posible mejorar el NNT con la distribución que se muestra en la Tabla 10C, que proporciona cinco aminoácidos (SER, LEU, HIS, VAL, ASP) en cantidades casi iguales. Ocho aminoácidos adicionales (PHE, TYR, ILE, ASN, PRO, ALA, ARG, GLY) están presentes en una abundancia del 78% de la de SER. THR y CYS permanecen a la mitad de la abundancia de SER. Cuando se variega el ADN para

micro-proteínas unidas con puentes disulfuro, es deseable a menudo reducir la prevalencia de CYS. Esta distribución permite observar 13 aminoácidos a un alto nivel y no proporciona interrupciones; la distribución de fxS optimizada permite solamente 11 aminoácidos con alta prevalencia.

El codón NNG también puede optimizarse. La Tabla 10D muestra un codón aproximadamente optimizado ([ALA] ~[ARG]) NNG. Existen, en esta variegación, cuatro aminoácidos más favorecidos de la misma manera: LEU, ARG, ALA y GLU. Véase que existe un aminoácido ácido y uno básico en este conjunto. Existen también dos aminoácidos menos favorecidos de la misma manera: TRP y MET. La proporción de lfaa/mfaa es de 0,5258. Si este codón se repite seis veces, los péptidos compuestos totalmente por TRP y MET son el 2% de comunes con respecto a los péptidos compuestos totalmente por los aminoácidos más favorables. A esto se le denomina "la prevalencia de $(TRP/MET)^6$ en ADNvg NNG^6 optimizado".

Cuando se sintetiza ADNvg mediante el método del "frasco sucio", es deseable a veces usar solamente una cantidad limitada de mezclas. Una mezcla muy útil se denomina la "mezcla NNS optimizada" en la que se promedian las dos primeras posiciones de la mezcla fxS: $T_1 = 0{,}24$, $C_1 = 0{,}17$, $A_1 = 0{,}33$, $G_1 = 0{,}26$, la segunda posición es idéntica a la primera, $C_3 = G_3 = 0{,}5$. Esta distribución proporciona los aminoácidos ARG, SER, LEU,: GLY, VAL, THR, ASN y LYS a más del 5% más ALA, ASP, GLU, ILE, MET y TYR a más del 4%.

También es interesante un codón variegado de forma compleja adicional. Este codón es idéntico al codón NNT optimizado en las primeras dos posiciones y tiene T:G::90:10 en la tercera posición. Este codón proporciona trece aminoácidos (ALA, ILE, ARG, SER, ASP, LEU, VAL, PHE, ASN, GLY, PRO, TYR y HIS) a más del 5,5%. THR al 4,3% y CYS al 3,9% son más comunes que los LFAA de NNK (3,125%). Los cinco aminoácidos restantes están presentes a menos del 1%. Este codón tiene la característica de que todos los aminoácidos están presentes; las secuencias que tienen más de dos de los aminoácidos de baja abundancia son raras. Cuando se aísla un SBD usando este codón, puede asegurarse de forma razonable que los primeros 13 aminoácidos se ensayaron en cada posición. Puede usarse un codón similar, en base al NNG optimizado.

La Tabla 10E muestra algunas propiedades de un codón NNS (o NNK) no optimizado. Obsérvese que existen tres aminoácidos más favorecidos de forma equivalente: ARG, LEU y SER. También existen *doce* aminoácidos *menos favorecidos* de forma equivalente: PHE, ILE, MET, TYR, HIS, GLN, ASN, LYS, ASP, GLU, CYS y TRP. Cinco aminoácidos (PRO, THR, ALA, VAL, GLY) están en el medio. Véase que una repetición de seis veces de NNS proporciona secuencias compuestas por los aminoácidos [PHE, ILE, MET, TYR, HIS, GLN, ASN, LYS, ASP, GLU, CYS y TRP] a solamente el ~0,1% de las secuencias compuestas por [ARG, LEU y SER]. Esto no solamente es ~20 veces inferior que la prevalencia de $(TRP/MET)^6$ en el ADNvg de NNG^6 optimizado, sino que esta prevalencia inferior se aplica a *doce* aminoácidos.

Mutagénesis Difusa

Puede aplicarse mutagénesis difusa a cualquier parte de la proteína en cualquier momento, pero es más adecuado cuando se ha establecido alguna unión a la diana. La mutagénesis difusa puede realizarse fijando cada uno de los nucleótidos puros activados para la síntesis de ADN (por ejemplo, nucleótido-fosforamiditas) con una pequeña cantidad de uno o más de los otros nucleótidos activados. Preferiblemente, el nivel de fijación se establece de modo que solamente un pequeño porcentaje (del 1% al 0,0001%, por ejemplo) del producto final contendrá la secuencia de ADN inicial. Esto asegurará que se produzcan muchas mutaciones sencillas, dobles, triples y superiores, pero que la recuperación de la secuencia básica será un resultado posible.

III.D. *Consideraciones Especiales en Relación con la Variegación de Micro-Proteínas con Cisteínas Esenciales*

Varios de los codones variegados simples o complejos preferidos codifican un conjunto de aminoácidos que incluye cisteína. Esto significa que algunos de los dominios de unión codificados presentarán una o más cisteínas además de las cisteínas unidas por puentes disulfuro invariables. Por ejemplo, en cada posición codificada por NNT, existe una posibilidad entre dieciséis de obtener una cisteína. Si se modifican seis codones de este modo, la fracción de dominio que contiene cisteínas adicionales es de 0,33. Cantidades extrañas de cisteína pueden conducir a complicaciones, véase Perry y Wetzel (PERR84). Por otro lado, muchas proteínas que contienen puentes disulfuro contienen cisteínas que no forman puentes disulfuro, por ejemplo, *tripsina*. La posibilidad de cisteínas no emparejadas puede tratarse de varias maneras:

En primer lugar, la población de fagos variegados puede pasarse por un reactivo inmovilizado que se une fuertemente a los tioles libres, tales como SulfoLink (número de catálogo 44895 H de Pierce Chemical Company, Rockford, Illinois, 61105). Otro producto de Pierce es TNB-Thiol Agarose (código de catálogo 20409 H). Para este fin, BioRad comercializa el Affi-Gel 401 (catálogo 153-4599).

En segundo lugar, puede usarse una variegación que excluye las cisteínas, tal como:

NHT que proporciona [F,S,Y,L,P,H,I,T,N,V,A,D].

VNS que proporciona $[L^2,P^2,H,Q,R^3,I,M,T^2,N,K,S,V^2,A^2,E,D,G^2]$.

NNG que proporciona [L^2,S,W,P,O,R^2,M,T,K,R,V,A,E,G, detención],

SNT que proporciona [L,P,H,R,V,A,D,G],

RNG que proporciona [M,T,K,R,V,A,E,G],

RMG que proporciona [T,K,A,E],

VNT que proporciona [L,P,H,R,I,T;N,S,V,A,D,G], o

RRS que proporciona [N,S,K,R,D,E,G^2].

Sin embargo, cada uno de estos esquemas tiene una o más de las desventajas, en relación con NNT: a) se permiten menos aminoácidos, b) los aminoácidos no se proporcionan de manera uniforme, c) los aminoácidos ácidos y básicos no son igualmente probables, o d) se producen codones de detención. Sin embargo, NNG, NHT y VNT son casi tan útiles como NNT. NNG codifica 13 aminoácidos diferentes y una señal de detención. Solamente dos aminoácidos aparecen dos veces en la mezcla de 16 veces.

En tercer lugar, puede enriquecerse la población para la unión a la diana preseleccionada, y evaluar las secuencias seleccionadas *post hoc* para cisteínas extra. Aquellos que contienen más cisteínas que las cisteínas proporcionadas para la limitación conformacional pueden usarse perfectamente. Es posible que se produzca un puente disulfuro diferente del designado. Esto no significa que el dominio de unión definido por la secuencia de ADN aislada sea de ninguna manera inadecuado. La idoneidad del dominio aislado se determina mejor mediante la evaluación química y bioquímica de péptidos sintetizados químicamente.

Por último, pueden bloquearse los tioles libres con reactivos, tales como reactivo de Ellman, yodo acetato o yoduro de metilo, que se unen específicamente a los tioles libres y que no reaccionan con los puentes disulfuro, y después dejan el fago modificado en la población. Debe entenderse que el agente bloqueante puede alterar las propiedades de unión de la micro-proteína; de este modo, pueden usarse diversos agentes bloqueantes si se espera que aparezcan diferentes dominios de unión. La población variegada de paquetes genéticos con el tiol bloqueado se fracciona para la unión. Si la secuencia de ADN de las micro-proteínas de unión aisladas contiene una cantidad impar de cisteínas, se usan medios sintéticos para preparar micro-proteínas que tienen cada puente posible y en las cuales el tiol impar se bloquea de forma apropiada. Nishiuchi (NISHB2, NISH86 y los trabajos mencionados en esos documentos) describen métodos para sintetizar péptidos que contienen una pluralidad de cisteína de modo que cada tiol se proteja con un tipo diferente de grupo bloqueante. Estos grupos pueden retirarse de forma selectiva de modo que pueda controlarse el emparejamiento disulfuro. Se prevé usar dicho esquema con la alteración de que un tiol permanece bloqueado, o se desbloquea y después se bloquea de nuevo con un reactivo diferente.

III.E. *Planificación de la Segunda y Posteriores Rondas de Variegación*

El método permite una acumulación eficaz de información referente a la secuencia de aminoácidos de un dominio de unión que tiene una gran afinidad por una diana predeterminada. Aunque puede obtenerse un dominio de unión muy útil a partir de una única ronda de variegación y de enriquecimiento de afinidad, se espera que sean necesarias múltiples rondas para alcanzar la afinidad y especificidad más altas posibles.

Si la primera ronda de variegación da como resultado alguna unión a la diana, pero la afinidad por la diana sigue siendo muy baja, puede alcanzarse una mejora adicional mediante la variegación de los SBD. El proceso puede ser progresivo, *es decir*, cada ciclo de variegación produce un punto de partida mejor para el siguiente ciclo de variegación de lo que produjo el ciclo anterior. El ajuste del nivel de variegación de modo que la secuencia *ppbd* y las muchas secuencias relacionadas con la secuencia *ppbd* estén presentes en cantidades detectables asegura que el proceso sea progresivo.

Si el nivel de variegación es lo suficientemente alto para que la secuencia *ppbd* esté presente en tan niveles bajos como para que exista una posibilidad apreciable de que ningún transformante presente el PPBD, entonces el mejor SBD de la siguiente ronda *podría* ser *peor* que el PPBD. A un nivel de variegación excesivamente alto, cada ronda de mutagénesis es independiente de las rondas previas y no se asegura la progresividad. Este enfoque puede conducir a proteínas de unión valiosas, pero la repetición de experimentos con este nivel de variegación no producirá resultados progresivos. No se prefiere la variación excesiva.

La progresividad no es una propiedad de todo o nada. Mientras que la mayoría de la información obtenida de ciclos de variegación anteriores se conserve y se produzcan muchas superficies diferentes que están relacionadas con el PPBD, el proceso será progresivo.

Si el nivel de variegación en el ciclo de variegación anterior se seleccionó correctamente, entonces los aminoácidos seleccionados para que estén en los restos recién modificados son los que mejor se determinaron. El entorno de otros restos ha cambiado, de modo que es apropiado para modificarlos de nuevo. Puesto que a menudo existen más residuos de interés que pueden modificarse simultáneamente, podemos continuar tomando restos que no se hayan modificado (prioridad más alta) o que no se hayan modificado durante uno o más ciclos.

El uso de codones variegados NNT o NNG conduce al muestreo muy eficaz de bibliotecas muy variegadas puesto que la proporción de (secuencias de aminoácidos diferentes)/(secuencias de ADN diferentes) es mucho más cercana a la unidad de lo que es para NNK o incluso para el codón vg optimizado (fxS). Sin embargo, se omiten unos pocos aminoácidos en cada caso. NNT y NNG permiten la obtención de miembros de todas las clases importantes de aminoácidos: hidrófobos, hidrófilos, ácidos, básicos, neutros hidrófilos, pequeños y grandes. Después de seleccionar un dominio de unión, puede ser deseable una variegación y selección posterior para alcanzar una mayor afinidad o especificidad. Durante esta segunda variegación, pueden investigarse las posibilidades de aminoácidos ignorados por la variegación anterior.

Unos pocos ejemplos pueden ser de ayuda. Se supone que se obtiene PRO usando NNT. Este aminoácido está disponible con NNT o NNG. Podemos estar razonablemente seguros de que PRO es el mejor aminoácido del conjunto [PRO, LEU, VAL, THR, ALA, ARG, GLY, PHE, TYR, CYS, HIS, ILE, ASN, ASP, SER]. A continuación podemos probar un conjunto que incluye [PRO, TRP, GLN, MET, LYS, GLU]. El conjunto permitido por NNG es el conjunto preferido.

¿Y si obtuviéramos en su lugar HIS? La Histidina es aromática y claramente hidrófoba, y puede formar puentes de hidrógeno con y a partir del anillo imidazol. El triptófano es hidrófobo y aromático y puede donar un hidrógeno a un receptor adecuado y estaba excluido por el codón NNT. La Metionina también se excluyó y es hidrófoba. De este modo, una ruta preferida es usar el codón variegado HDS que proporciona [HIS, GLN, ASN, LYS, TYR, CYS, TRP, ARG, SER, GLY, <detención>].

Si la primera ronda de variegación no tiene ningún éxito, debe usarse un patrón de variegación diferente. Por ejemplo, si puede definirse más de un conjunto de interacción con un dominio, los restos modificados en la siguiente ronda de variegación deben pertenecer a un conjunto diferente del que se sondeó en la variegación inicial. Si se producen fallos repetidos, puede cambiarse a un IPBD diferente.

IV. *Estrategia de presentación: presentación de dominios de unión extraños sobre la superficie de un "paquete genético"*

IV.A. *Requisitos Generales para los Paquetes Genéticos*

Para obtener la presentación de una multitud de dominios de unión potenciales diferentes aunque relacionados, los solicitante generan una población heterogénea de paquetes genéticos replicables, cada uno de los cuales comprende un gen híbrido que incluye una primera secuencia de ADN que codifica un dominio de unión potencial para la diana de interés y una segunda secuencia de ADN que codifica un medio de presentación, tal como una proteína de la superficie externa nativa para el paquete genético pero no asociada de forma nativa con el dominio de unión potencial (o con el dominio de unión parental con el que está relacionado) que provoca que el paquete genético presente la proteína quimérica correspondiente (o una forma procesada de la misma) en su superficie externa.

El componente de una población que muestra las propiedades de unión deseadas puede ser bastante pequeño, por ejemplo, uno de 10^6 o menos. Una vez que este componente de la población se ha separado de los componentes no de unión, debe ser posible amplificarlo. El cultivo de células viables es la amplificación más poderosa de material genética que se conoce y se prefiere. También pueden amplificarse *in vitro* mensajes genéticos, por ejemplo mediante PCR, pero éste no es el método más preferido.

Preferiblemente, el GP puede: 1) alterarse genéticamente con una facilidad razonable para codificar un dominio de unión potencial, 2) mantenerse y amplificarse en cultivo, 3) manipularse para presentar el dominio de unión potencial donde pueda interaccionar con el material diana durante la separación por afinidad y 4) separarse por afinidad mientras conserva la información genética que codifica el dominio de unión presentado en forma recuperable. Preferiblemente, el GP sigue siendo viable después de la separación por afinidad. Los GP preferidos son células bacterianas vegetativas, esporas bacterianas y especialmente virus del ADN bacteriano. Pueden usarse células eucariotas y virus eucariotas como paquetes genéticos, pero no se prefieren.

Cuando el paquete genético es una célula bacteriana o un fago que está unido por el periplasma, el medio de presentación tiene dos componentes. El primer componente es una señal de secreción que dirige al producto de la expresión inicial hacia la membrana interna de la célula (una célula hospedadora cuando el paquete es un fago). Esta señal de secreción se retira por escisión mediante una señal peptidasa para proporcionar una proteína de unión potencial procesada y madura. El segundo componente es una señal de transporte de la superficie externa que dirige el paquete para que se una con la proteína procesada en su superficie externa. Esta señal de transporte de la superficie externa se obtiene a partir de una proteína de la superficie nativa del paquete genético.

Por ejemplo, en una realización preferida, el gen híbrido comprende un ADN que codifica un dominio de unión potencial unido de forma operativa a una secuencia de señal (*por ejemplo*, las secuencias de señal de los genes bacterianos *phoA* o *bla* en la secuencia de señal del *gen III* del fago M13) y al ADN que codifica una proteína de cubierta (*por ejemplo*, las proteínas del gen III o del gen VIII de M13) de un fago filamentoso (*por ejemplo*, M13). El producto de expresión se transporta hasta la membrana interna (bicapa lipídica) de la célula hospedadora, mientras tanto el péptido señal se retira por escisión para dejar una proteína híbrida procesada. El extremo C del componente similar a la proteína de cubierta de esta proteína híbrida está atrapado en la bicapa lipídica, de modo que la proteína híbrida

no escapa hacia el espacio periplásmico. (Esto es típico de la proteína de cubierta de tipo silvestre). A medida que el ADN de cadena sencilla de la partícula del fago naciente pasa hacia el espacio periplásmico, recoge la proteína de cubierta de tipo silvestre y la proteína híbrida de la capa lipídica. La proteína híbrida se envasa de este modo en la capa superficial del fago filamentoso, dejando el dominio de unión potencial expuesto en su superficie externa. (De este modo, el fago filamentoso, y no la célula bacteriana hospedadora, es el "paquete genético replicable" en esta realización).

Si es necesaria una señal de secreción para la presentación del dominio de unión potencial, en una realización especialmente preferida la célula bacteriana en la que se expresa el gen híbrido es de una cepa "de secreción permisiva".

Cuando el paquete genético es una espora bacteriana o un fago (tal como ϕX174 o λ) cuya cubierta se une de forma intracelular, no es necesaria una señal de secreción que dirija el producto de expresión hacia la membrana interna de la célula bacteriana hospedadora. En estos casos, el medio de presentación es simplemente la señal de transporte de la superficie externa, típicamente un derivado de una proteína de cubierta de una espora o fago.

Los OSP preferidos para varios GP se proporcionan en la Tabla 2. Las referencias de fusiones *osp-ipbd* en esta sección deben tomarse como aplicadas, *mutatis mutandis*, a fusiones *osp-pbd* y también a *osp-sbd*.

IV.B. *Fagos Para su Uso como GP*

Se prefieren fagos unidos en el periplasma (el IPBD es una micro-proteína unida con puentes disulfuro), ya que los IPBD no pueden plegarse dentro de una célula (estas proteínas pueden plegarse después de que el fago se libere de la célula). Se prefieren fagos unidos de forma intracelular cuando el IPBD necesita grupos prostéticos grandes o insolubles (tales como agrupaciones de Fe_4S_4), puesto que el IPBD no puede plegarse si se secreta ya que el grupo prostético no está en el periplasma.

Cuando se introduce variegación, múltiples infecciones podrían generar GP híbridos que tienen el gen para un PBD pero tienen al menos algunas copias de algún PBD diferente en sus superficies; es preferible minimizar esta posibilidad infectando las células con el fago en condiciones resultantes de una baja multiplicidad-de-infección (MOI).

Los bacteriófagos son candidatos excelentes para GP puesto que existe poca o ninguna actividad enzimática asociada con el fago maduro intacto, y puesto que los genes son inactivos fuera de un hospedador bacteriano, lo que hace a las partículas del fago maduro metabólicamente inertes.

Para un bacteriófago dado, el OSP preferido es normalmente uno que está presente en la superficie del fago en la mayor cantidad de copias. Sin embargo, un OSP tal como la proteína gIII de M13 (5 copias/fago) puede ser una excelente elección como OSP para provocar la presentación del PBD.

Se prefiere que se preserve el gen *osp* de tipo silvestre. Puede insertarse el fragmento del gen *ipbd* en una segunda copia del gen *osp* receptor o en un nuevo gen *osp* manipulado genéticamente. Se prefiere que el gen *osp-ipbd* se coloque bajo el control de un promotor regulado.

El usuario debe seleccionar un sitio en el gen OSP candidato para insertar un fragmento del gen *ipbd*. Las cubiertas de la mayoría de los bacteriófagos están altamente ordenadas. En dichos bacteriófagos, es importante conservar las proteínas de fusión OSP-IPBD manipuladas genéticamente cuyos restos del OSP parental interaccionarán con otras proteínas en el virión. Para gVIII de M13, se conserva preferiblemente toda la proteína madura, mientras que para gIII de M13, puede ser suficiente conservar los últimos 100 restos (BASS90) (o incluso menos). Dicha proteína gIII truncada se expresaría en paralelo con la proteína gIII completa, ya que se requiere la proteína gIII para la infectividad del fago.

Los fagos filamentosos, que incluyen M13, f1, fd, If1, Ike, Xf, Pf1 y Pf3, son de particular interés. La principal proteína de la cubierta la codifica el gen VIII. La proteína de cubierta del gen VIII madura de 50 aminoácidos se sintetiza como una precubierta de 73 aminoácidos (ITOK79). Los primeros 23 aminoácidos constituyen una señal de secuencia típica que provoca que el polipéptido naciente se inserte en la membrana interna de la célula.

Una peptidasa de señal de *E. coli* (SP-I) reconoce los aminoácidos 18, 21 y 23 y, en menor medida, el resto 22 y corta entre los restos 23 y 24 de la precubierta. (KUHN85a, KUHN85b, OLIV87). Después de la retirada de la secuencia señal, el extremo amino de la cubierta madura se sitúa en el lado del periplasma de la membrana interna; el extremo carboxilo está en el lado citoplásmico. Aproximadamente 3000 copias de la proteína de cubierta de 50 aminoácidos madura se asocian lado a lado en la membrana interna.

La secuencia del gen *VIII* se conoce, y la secuencia de aminoácidos puede codificarse en un gen sintético, usando el promotor *lacUV5* junto con el represor $Lacl^q$. El promotor *lacUV5* está inducido por IPTG. La proteína del gen VIII madura forma la cubierta alrededor del ADNss circular. La estructura 3D del virión f1 se conoce a resolución media; el extremo amino de la proteína del gen VIII está en la superficie del virión y es por lo tanto un sitio de unión preferido para el dominio de unión potencial. Se han realizado unas pocas modificaciones del gen *VIII* que se analizan a continuación. La estructura 2D de la proteína de cubierta de M13 está implícita en la estructura 3D. La proteína del gen VIII de M13 madura solamente tiene un dominio.

Se ha construido un gen tripartito que comprende:

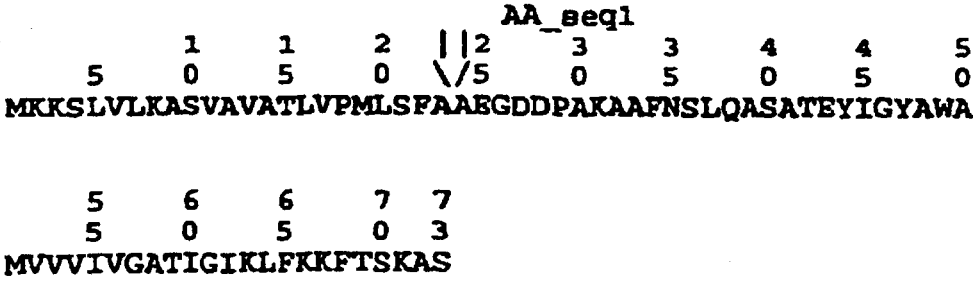
1) ADN que codifica una secuencia señal que dirige la secreción de las partes (2) y (3) a través de la membrana interna,

2) ADN que codifica la secuencia BPTI madura, y

3) ADN que codifica la proteína gVIII de M13.

Este gen provoca que BPTI aparezca en forma activa en la superficie del fago M13.

La secuencia de aminoácidos de la precubierta de M13 (SCHA78), denominada AA_seq1, es



El mejor sitio para insertar un dominio de proteína nueva en el citoplasma de M13 es después de A23 puesto que SP escinde la proteína de precubierta después de A23, como indica la flecha. Las proteínas que pueden secretarse aparecerán conectadas al citoplasma de M13 maduro en su extremo amino. Puesto que el extremo amino del CP de M13 maduro se sitúa en la superficie externa del virión, el dominio introducido se presentará en la parte externa del virión. El desconocimiento del mecanismo por el cual M13CP aparece en la bicapa lipídica genera la posibilidad de que la inserción directa del *bpti* en el gen *VIII* no produzca una proteína de fusión funcional. Puede ser necesario cambiar la secuencia señal de la fusión a, por ejemplo, la secuencia señal *phoA* (MKQSTIALA-LLPLLFTPVTKA...) (MARK91). Marks *et al*. (MARK86) demostraron que el péptido señal *phoA* podía dirigir el BPTI maduro al periplasma de *E. coli*.

Otro medio para presentar el IPBD es expresarlo en forma de un dominio de un gen quimérico que contiene parte o todo el gen *III*. Este gen codifica una de las proteínas de cubierta secundarias de M13. Los genes VI, VII y IX también codifican proteínas de cubierta secundarias. Cada una de estas proteínas secundarias se presenta en aproximadamente 5 copias por virión y se relaciona con la morfogénesis o la infección. Por el contrario, la proteína de cubierta principal se presenta en más de 2500 copias por virión. Las proteínas del gen VI, VII y IX se presentan en los extremos del virión. Estas tres proteínas no se procesan de forma post-traduccional (RASC86).

El ADN del fago circular de cadena sencilla se asocia con aproximadamente 5 copias de la proteína del gen III y después se extruye a través del parche de proteína de cubierta asociada a la membrana de tal manera que el ADN se encierra en una vaina helicoidal de proteína (WEBS78). El ADN no empareja las bases (lo que podría imponer graves restricciones al genoma del virus); en su lugar, las bases se intercalan entre sí independientemente de la secuencia.

Smith (SMIT85) y de la Cruz *et al*. (DELA88) han demostrado que las inserciones en el gen *III* provocan que aparezcan dominios de la nueva proteína en la superficie externa del virión. El gen de la mini-proteína puede fusionarse con el gen *III* en el sitio usado por Smith y por de la Cruz *et al*., en un codón correspondiente a otro dominio de frontera o a un bucle de la superficie de la proteína, o al extremo amino de la proteína madura.

Todos los trabajos publicados usan un vector que contiene un único gen *III* modificado de fd. De este modo, las cinco copias de gIII se modifican de forma idéntica. El gen *III* es bastante grande (1272 p.b. o aproximadamente el 20% del genoma del fago) y se desconoce si un duplicado del gen completo puede insertarse de forma estable en el fago. Además, las cinco copias de la proteína gIII están en un extremo del virión. Cuando las moléculas diana bivalentes (tales como anticuerpos) se unen a un fago pentavalente, el complejo resultante puede ser irreversible. La unión irreversible del GP a la diana interfiere en gran medida con un enriquecimiento de afinidad de los GP que tienen las secuencias genéticas que codifican el nuevo polipéptido que tiene la mayor afinidad por la diana.

Para reducir la posibilidad de formación de complejos irreversibles, podemos usar un segundo gen sintético que codifica las partes carboxilo terminales de *III*; las partes carboxilo terminales de la proteína del gen *III* provocan que se una al fago. Por ejemplo, los 29 restos finales (empezando con la arginina especificada por el codón 398) pueden

ser suficientes para provocar que una proteína de fusión se una al fago. Como alternativa, puede incluirse el dominio globular final de la proteína gIII madura, es decir, los 15 a 160 aminoácidos finales del gen III (BASS90). Por ejemplo, puede formarse por manipulación genética un gen compuesto por (de 5' a 3'):

1) un promotor (preferiblemente regulado),

2) un sitio de unión al ribosoma,

3) un codón de iniciación,

4) un péptido señal funcional que dirige la secreción de las partes (5) y (6) a través del a membrana interna,

5) ADN que codifica un IPBD,

6) ADN que codifica los restos 275 a 424 de la proteína M13 gIII,

7) un codón de detención de la traducción, y

8) (opcionalmente) una señal de detención de la transcripción.

Se deja el gen *III* de tipo silvestre, de modo que alguna proteína del gen III sin alterar estará presente. Como alternativa, puede usarse la proteína del gen VIII como OSP y regular la fusión *osp::ipbd* de modo que solamente unas pocas copias de la proteína de fusión aparezcan en el fago.

Las proteínas del gen VI, VII y IX de M13 no se procesan después de la traducción. La vía mediante la cual estas proteínas se unen al fago no se ha descrito. Estas proteínas son necesarias para la morfogénesis y la infectividad normales del fago. No se conoce si estas moléculas (Proteína del gen VI, proteína del gen VII y proteína del gen IX) se unen al fago: a) desde el citoplasma, b) desde el periplasma, o c) desde dentro de la bicapa lipídica. Puede usarse cualquiera de estas proteínas para introducir un IPBD en la superficie del fago mediante una de las construcciones:

1) *ipbd::pmcp*,

2) *pmcp::ipbd*,

3) *señal::ipbd::pmcp*, y

4) *señal::pmcp::ipbd*.

donde *ipbd* representa el ADN que codifica la expresión del domino de unión potencial inicial; *pmcp* representa el ADN que codifica una de las proteínas de cubierta secundaria del fago, VI, VII y IX; *señal* representa un péptido señal de secreción funcional, tal como la señal *phoA* (MKQSTIALALLPLLFTPVTYA); y "::" representa una fusión genética en marco. Las fusiones indicadas se colocan cadena debajo de un promotor conocido, preferiblemente un promotor regulado tal como *lacUV5*, *tac* o *trp*. Las fusiones (1) y (2) son apropiadas cuando la proteína de cubierta secundaria se une al fago desde el citoplasma o mediante una inserción autónoma en la bicapa lipídica. La fusión (1) es apropiada si el extremo amino de la proteína de cubierta secundaria está libre y la (2) es apropiada si el extremo carboxilo está libre. Las fusiones (3) y (4) son apropiadas si la proteína de cubierta secundaria se une al fago desde el periplasma o desde dentro de la bicapa lipídica. La fusión (3) es apropiada si el extremo amino de la proteína de cubierta secundaria está libre y la (4) es apropiada si el extremo carboxilo está libre.

Pueden prepararse construcciones similares con otros fagos filamentosos. Pf3 es un fago filamentoso bien conocido que infecta las células de *Pseudomonas aerugenosa* que aloja un plásmido IncP-1. La proteína de cubierta principal de PF3 no es habitual puesto que no tiene péptido señal para dirigir esta secreción. La secuencia tiene restos cargados ASP_7, ARG_{37}, LYS_{40} y PRE_{44}-COO^-, lo que es coherente con la exposición del extremo amino. De este modo, para provocar la aparición de IPBD en la superficie de Pf3, se construye un gen tripartito que comprende:

1) una secuencia señal, que se sabe que provoca la secreción en *P. aerugenosa* (que preferiblemente se sabe que provoca la secreción de IPBD) fusionado en marco a,

2) un fragmento génico que codifica la secuencia de IPBD, fusionado en marco a,

3) un ADN que codifica la proteína de cubierta de Pf3 madura.

Opcionalmente, se introduce ADN que codifica un enlazador flexible de uno a 10 aminoácidos y/o aminoácidos que forman un sitio de reconocimiento para una proteasa específica (por ejemplo, Factor Xa) entre el fragmento del gen *ipbd* y el gen de la *proteína de cubierta* de Pf3. Este gen tripartito se introduce en Pf3 de modo que no interfiera con la expresión de cualquier gen de Pf3. Para reducir la posibilidad de recombinación genética, la parte (3) se diseña para que tenga numerosas mutaciones silenciosas en relación con el gen de tipo silvestre. Una vez que la secuencia señal se ha retirado por escisión, el IPBD está en el periplasma y la proteína de cubierta madura actúa como un anclaje

y una señal de unión del fago. No importa que esta proteína de fusión descanse sobre la bicapa lipídica mediante una ruta diferente de la ruta seguida por la proteína de cubierta de tipo silvestre.

Como se describe en el documento WO 90/02809, puede usarse otro fago, tal como el bacteriófago ϕX174, fago de ADN grande tal como λ o T4 e incluso fagos de ARN, con adaptaciones y modificaciones adecuadas como GP.

IV.C. *Células Bacterianas como Paquetes Genéticos*

Puede seleccionarse cualquier cepa bacteriana bien caracterizada que (1) pueda cultivarse, (2) pueda manipularse genéticamente para presentar PBD en su superficie, y (3) sea compatible con la selección de afinidad.

Entre las células bacterianas, los paquetes genéticos preferidos son *Salmonella typhimurium, Bacillus subtilis, Pseudomonas aeruginosa, Vibrio cholerae, Klebsiella pneumonia, Neisseria gonorrhoeae, Neisseria meningitidis, Bacterioides nodosus, Moraxella bovis* y especialmente *Escherichia coli*. La mini-proteína de unión potencial puede expresarse en forma de un inserto en una proteína de la superficie externa bacteriana quimérica (OSP). Todas las bacterias muestran proteínas en sus superficies externas. *E. coli* es el GP bacteriano preferido y para éste LamB es una OSP preferida.

Aunque la mayoría de las proteínas bacterianas permanecen en el citoplasma, otras se transportan al espacio periplasmático (que está entre la membrana plasmática y la pared celular de las bacterias gram negativas) o se transportan o se anclan a la superficie externa de la célula. Otras más se exportan (se secretan) al medio que rodea a la célula. Estas características de una proteína que reconoce una célula y que provocan que se transporten fuera del citoplasma y se presenten en la superficie celular se denominarán "señales de transporte a la superficie externa".

Las bacterias gram negativas tienen proteínas de la membrana externa (OMP) que forman un subconjunto de OSP. Muchas OMP se prolongan a lo largo de la membrana una o más veces. Las señales que provocan que las OMP se sitúen en la membrana externa están codificadas en la secuencia de aminoácidos de la proteína madura. Las proteínas de la membrana externa de las bacterias se expresan inicialmente en una secuencia de aminoácidos precursora de la proteína madura. Las proteínas de la membrana externa de las bacterias se expresan inicialmente en una forma precursora incluyendo una llamada péptido señal. La proteína precursora se transporta hacia la membrana interna y el resto del péptido señal se extruye al espacio periplasmático. Allí, se retira por escisión mediante una "peptidasa de señal" y el resto de la proteína "madura" puede entrar ahora en el periplasma. Una vez que está allí, otros mecanismos celulares reconocen las estructuras en la proteína madura que indican cuál es su lugar adecuado en la membrana externa y la transportan hacia ese punto.

Se sabe bien que el ADN que codifica el péptido líder o señal de una proteína puede unirse a la secuencia de ADN que codifica otra proteína, proteína X, para formar un gen quimérico cuya expresión provoca que la proteína X aparezca libre en el periplasma. El uso de cepas bacterianas permisivas con la exportación (LISS85, STAD89) aumenta la probabilidad de que una fusión de señal-secuencia dirija la proteína deseada a la superficie celular.

Las proteínas de fusión OSP-IPBD no necesitan cumplir un papel estructural en las membranas externas de bacterias gram negativas ya que partes de las membranas externas no están altamente ordenadas. Para grandes OSP es probable que haya uno o más sitios en los que las *osp* pueden truncarse y fusionarse con *ipbd* de modo que las células que expresan la fusión presentarán IPBD en la superficie celular. Las fusiones de fragmentos de genes *omp* con fragmentos de un gen *x* han conducido a la aparición de X en la membrana externa (CHAR88b,c, BENS84, CLEM81). Cuando se han realizado dichas fusiones, puede diseñarse un gen *osp-ipbd* sustituyendo *x* por *ipbd* en la secuencia de ADN. De otro modo, se busca preferiblemente una fusión exitosa de OMP-IPBD fusionando fragmentos del mejor *omp* con un *ipbd*, expresando el gen fusionado y ensayando los GP resultantes para la presentación del fenotipo IPBD. Se usan los datos disponibles acerca de la OMP para tomar el punto o puntos de fusión entre *omp* e *ipbd* para maximizar las posibilidades de que se presente ese IPBD (el ADN espaciador que codifica enlazadores flexibles, preparado *por ejemplo* con GLY, SER y ASN, puede colocarse entre los fragmentos obtenidos de *osp* e *ipbd* para facilitar la presentación). Como alternativa, se trunca *osp* en varios sitios o de manera que se produzcan fragmentos de *osp* de longitud variable y se fusionan los fragmentos de *osp* con *ipbd*; las células que expresan la fusión se exploran o se seleccionan las que presentan IPBD en la superficie celular. Freudl *et al*. (FREU89) han demostrado que los fragmentos de OSP (tales como OmpA) por encima de cierto tamaño se incorporan en la membrana externa. Una alternativa adicional es incluir segmentos cortos de ADN aleatorio en la fusión de fragmentos de *omp* con *ipbd* y después explorar o seleccionar la población variegada resultante para miembros que muestran el fenotipo de presentación de IPBD.

En *E. coli*, la proteína LamB es una OSP bien conocida y puede usarse. La Lamb de *E. coli* se ha expresado de forma funcional en *S. typhimurium, V. cholerae* y *K. pneumonia*, de modo que podría presentarse una población de PBD en cualquiera de estas especies en forma de una fusión con LamB de *E. coli*. *K. pneumonia* expresa una maltoporina similar a LamB (WEHM89) que también podría usarse. En *P. aeruginosa*, puede usarse la proteína D1 (un homólogo de LamB) (TRIA88).

LamB se transporta a la membrana externa si está presente una secuencia N-terminal funcional. Además, se requieren los primeros 49 aminoácidos de la secuencia madura para el transporte con éxito (BENS84). Al igual que con otras OSP, LamB de *E. coli* se sintetiza con una secuencia señal típica que posteriormente se retira. Se ha detectado homología entre partes de la proteína LamB y las otras proteínas de la membrana externa OmpC, OmpF y PhoE

(NIKA84), incluyendo homología entre los aminoácidos 39-49 de LamB y las secuencias de las otras proteínas. Estas subsecuencias pueden marcar las proteínas para el transporte a la membrana externa.

Se conoce la secuencia de aminoácidos de LamB (CLEM81) y se ha desarrollado un modelo de cómo ésta se ancla a sí misma a la membrana externa (Revisado por, entre otros, BENZ88b). La situación de sus dominios de maltosa y de unión al fago también se conoce (HEIN88). Usando esta información, pueden identificarse varias estrategias mediante las cuales puede incorporarse un inserto de PBD en LamB para proporcionar una OSP quimérica que presenta el PBD en la membrana externa bacteriana.

Cuando los PBD deben presentarse mediante una proteína transmembrana quimérica como LamB, el PBD podría insertarse en un bucle que se encuentra normalmente en la superficie de la célula (*cp*. BECK83, MANO86). Como alternativa, puede fusionarse un segmento 5' del gen osp en el fragmento génico *ipbd*; el punto de fusión se toma para que corresponda con un bucle expuesto de la superficie de la OSP y se omiten las porciones terminales carboxilo de la OSP. En LamB, se ha descubierto que pueden insertarse hasta 60 aminoácidos (CHAR88b,c) con la resultante presentación del epítopo extraño; las características estructurales de OmpC, OmpA, OmpF y PhoE son tan similares que se espera un comportamiento similar de estas proteínas.

Debe observarse que aunque LamB puede caracterizarse como proteína de unión, se usa en la presente invención para proporcionar un OSTS; sus dominios de unión no se variegan.

Pueden usarse otras proteínas de la superficie externa bacteriana tales como OmpA, OmpC, OmpF, PhoE y pilina en lugar de LamB y sus homólogos. OmpA es de particular interés puesto que es muy abundante y porque se conocen homólogos en una amplia diversidad de especies bacterianas gram negativas. Baker *et al*. (BAKE87) revisan el conjunto de proteínas en la membrana externa de *E. coli* y citan un modelo topológico de OmpA (VOGE86) que predice que los restos 19-32, 62-73, 105-118 y 147-158 están expuestos en la superficie celular. La inserción de un fragmento que codifica *ipbd* en aproximadamente el codón 111 o en aproximadamente el codón 152 es probable que provoque que el IPBD se presente en la superficie celular. En relación con OmpA, véase también MACI88 y MANO88. La Proteína F de Porina de *Pseudomonas aeruginosa* se ha clonado y tiene una homología de secuencia con OmpA de *E. coli* (DUCH88). Aunque esta homología no es suficiente para permitir una predicción de los restos expuestos en superficie de la proteína F porina, los métodos usados para determinar el modelo topológico de OmpA pueden aplicarse a la Proteína F de Porina. Los trabajos relacionados con el uso de OmpA como una OSP incluyen BECK80 y MACI88.

Misra y Benson (MISR88a, MISR88b) describen un modelo topológico de OmpC de *E. coli* que predice que, entre otros, los restos GLY_{164} y LEU_{250} están expuestos en la superficie celular. De este modo, la inserción de un fragmento del gen *ipbd* a aproximadamente el codón 164 o en aproximadamente el codón 250 del gen *ompC de E. coli* o en los codones correspondientes en el gen *ompC de S. typhimurium* es probable que cause la aparición del IPBD en la superficie celular. Pueden usarse los genes *ompC* de otras especies bacterianas. Otros trabajos relacionados con OmpC incluyen CATR87 y CLIC88.

La OmpF de *E. coli* es una OSP muy abundante, $\geq 10^4$ copias/célula. Pages *et al*. (PAGE90) han publicado un modelo de OmpF que indica siete segmentos expuestos en la superficie. La fusión de un fragmento del gen de *ipbd*, en forma de inserto o para sustituir la parte 3' de *ompF*, en una de las regiones indicadas, es probable que produzca un gen *ompf::ipbd* funcional cuya expresión conduce a la presentación de IPBD en la superficie celular. En particular, la fusión en aproximadamente el codón 111, 177, 217 ó 245 debe conducir a un gen *ompF::ipbd* funcional. En relación con OmpF, véase también REID88b, PAGE88, BENS88, TOMM82 y SODE85.

Las proteínas de los pilus son de particular interés puesto que las células con pili expresan muchas copias de estas proteínas y puesto que varias especies (*N. gonorrhoeae, P. aeruginosa, Miraxella bovis, Bacteroides nodosus* y *E. coli*) expresan pilinas relacionadas. Getzoff y colaboradores (GETZ88, PARG87, SOME85) han construido un modelo del pilus de gonococos que predicen que la proteína formará un haz de cuatro hélices que tiene similitudes estructurales con la proteína del virus del mosaico del tabaco y con miohemeritrina. En este modelo, los extremos amino y carboxilo de la proteína están expuestos. El extremo amino está metilado. Elleman (ELLE88) ha revisado pilinas de *Bacteroides nodosus* y otras especies y las diferencias de serotipo pueden relacionarse con diferencias en la proteína pilina y la variación más importante se produce en la región C-terminal. Las porciones amino terminales de la proteína pilina están muy conservadas. Jennings *et al*. (JENN89) han insertado un fragmento de virus de la fiebre aftosa (restos 154-159) en la proteína fímbrica de tipo 4 de *B. nodosus* que es muy homóloga con la pilina de gonococos. Descubrieron que la expresión de la fusión 3'-terminal en *P. aeruginosa* conducía a una cepa viable que fabrica cantidades detectables de la proteína de fusión. Jennings *et al*. no modificaron el epítopo extraño ni sugirieron ninguna modificación. Insertaron un enlazador GLY-GLY entre el último resto de pilina y el primer resto del epítopo extraño para proporcionar un "enlazador flexible". De este modo, un sitio preferido para unir un IPBD es el extremo carboxilo. También pueden usarse los bucles expuestos del haz, aunque las fusiones internas particulares ensayadas por Jennings *et al*. (JENN89) parecían ser letales para *P. aeruginosa*. En relación con la pilina, véase también MCKE85 y ORND85.

Judd (JUDD86, JUDD85) ha investigado la proteína IA de *N. gonorrhoeae* y ha descubierto que el extremo amino está expuesto; de este modo, podría unirse un IPBD en o cerca del extremo amino de la P.IA madura como medio para presentar el IPBD en la superficie de *N. gonorrhoeae*.

Se ha descrito un modelo de la topología de PhoE de *E. coli* por van der Ley *et al*. (VAND86). Este modelo predice ocho bucles que están expuestos; es probable que la inserción de un IPBD en uno de estos bucles conduzca a la presentación del IPBD en la superficie de la célula. Los restos 158, 201, 238 y 275 son situaciones preferidas para la inserción de un IPBD.

Otros OSP que podían usarse incluyen BtuB, FepA, FhuA, lutA, FecA y FhuE de *E. coli* (GUDM89) que son receptores de nutrientes que se encuentran normalmente con poca abundancia. Los genes de todas estas proteínas se han secuenciado, pero aún no están disponibles los modelos topológicos. Gudmunsodttir *et al*. (GUDM89) han comenzado la construcción de dicho modelo para BtuB y FepA, mostrando que ciertos restos de BtuB se encaran con el periplasma y determinando la funcionalidad de diversas fusiones de BtuB::FepA. Carmel *et al*. (CARM90) han presentado un trabajo de naturaleza similar para FhuA. Todas las especies de Neisseria expresan proteínas de la superficie externa para el transporte de hierro que se han identificado y, en muchos casos, clonado. Véase también MORS87 Y MORS88.

Muchas bacterias gram negativas expresan una o más fosfolipasas. La fosfolipasa A de *E. coli*, producto del gen *pldA*, se ha clonado y secuenciado por Geus *et al*. (DEGE84). Éstos descubrieron que la proteína aparece en la superficie celular sin ningún procesamiento post-traduccional. Puede unirse un fragmento de gen ipbd en el extremo o insertarse en posiciones predichas para codificar bucles en la proteína. Esta fosfolipasa A llega a la superficie externa sin la retirada de una secuencia señal, lo que no demuestra que una proteína de fusión PldA::IPBD también seguirá esta ruta. De este modo, podemos provocar la secreción de una fusión PldA::IPBD o IPBD::PldA en el periplasma mediante la adición de una secuencia señal apropiada. De este modo, además de la fusión binaria simple de un fragmento de *ipbd* a un extremo de *pldA*, deben ensayarse las construcciones:

1) *ss::ipbd::pldA*

2) *ss::pldA::ipbd.*

Una vez que la proteína PldA::IPBD está libre en el periplasma, ya no recuerda cómo llegó hasta allí y las características estructurales de la PldA que provocaron que se situara en la superficie externa dirigirán la fusión al mismo destino.

IV.D. *Esporas Bacterianas Como Paquetes Genéticos*

Las esporas bacterianas tienen propiedades deseables como candidatos a GP. Las esporas son mucho más resistentes que las células bacterianas vegetativas o los fagos a los agentes químicos y físicos, y por lo tanto permiten el uso de una gran diversidad de condiciones de selección de afinidad. Además, las esporas de *Bacillus* ni metabolizan de forma activa ni alteran las proteínas en su superficie. Las esporas de *Bacillus* y más específicamente las esporas de *B. subtilis* son por lo tanto los GP de esporas preferidos. Como se analiza con más detalle en el documento WO 90/02809, puede introducirse un dominio de unión extraño en una proteína de la superficie externa como el codificado por los genes cotC y cotD de *B. subtilis*.

Es preferible generalmente usar como paquete genético una célula, espora o virus para el que ya se ha identificado una proteína de la membrana externa que pueda manipularse genéticamente para que presente un IPBD. Sin embargo, como se explica en el documento WO 90/02809, la presente invención no se limita a dichos paquetes genéticos ya que puede generarse otra señal de transporte de la superficie mediante técnicas de variegación y selección.

V.E *Construcción Genética y Consideraciones Sobre la Expresión*

El gen (i)*pbd-osp* puede ser: a) completamente sintético, b) un compuesto de ADN sintético y natural, o c) un compuesto de fragmentos de ADN naturales. El punto importante es que el segmento de *pbd* pueda variegarse fácilmente para codificar una familia diversa y multitudinaria de PBD como se ha descrito anteriormente. Se prefiere un segmento de *ipbd* sintético ya que permite un mayor control sobre la colocación de sitios de restricción. Para secuenciar se necesitan cebadores complementarios a las regiones colindantes al gen *osp-ipbd* en su extremo 3' y a las partes del gen *osp-ipbd* que no deben modificarse.

Las secuencias de las partes reguladoras del gen se toman de las secuencias de elementos reguladores naturales: a) promotores, b) secuencias Shine-Dalgarno, y c) terminadores de la transcripción. Los elementos reguladores también podían diseñarse a partir del conocimiento de secuencias consenso de regiones reguladoras naturales. Las secuencias de estos elementos reguladores están conectadas con las regiones codificantes; también se insertan sitios de restricción en o adyacentes a las regiones reguladoras para permitir una manipulación conveniente.

La función esencial de la separación por afinidad es separar los GP que tienen PBD (obtenidos de IPBD) que tienen alta afinidad por la diana de los GP que tienen PBD que tienen baja afinidad por la diana. Si el volumen de elución de un GP depende de la cantidad de PBD en la superficie del GP, entonces un GP que tiene muchos PBD con baja afinidad, GP(PBD_w) puede co-eluirse junto con un GP que tenga menos PBD con alta afinidad, GP (PBD_S). La regulación del gen *osp-pbd* es preferiblemente tal que la mayoría de los paquetes presentan suficiente PBD para realizar una buena separación de acuerdo con la afinidad. El uso de un promotor regulable para contro-

lar el nivel de expresión del *osp-pbd* permite un ajuste preciso del comportamiento cromatográfico de la población variegada.

La inducción de la síntesis de genes manipulados genéticamente en células bacterianas vegetativas se ha realizado a través del uso de promotores regulados tales como *lacUV5*, *trpP* o *tac* (MANI82). Los factores que regulan la cantidad de proteína sintetizada se conocen lo suficientemente bien como para que pueda producirse actualmente una amplia diversidad de proteínas heterólogas en *E. coli*, *B. subtilis* y otras células hospedadoras en al menos cantidades moderadas (BETT88). Preferiblemente, el promotor par el gen *osp-ipbd* se somete a regulación mediante un pequeño inductor químico, por ejemplo, el promotor *lac* y el promotor híbrido *trp-lac* (tac) son regulables con ispropil tiogalactósido (IPTG). No es necesario que el promotor del gen constructor provenga de un gen *osp* natural; puede usarse cualquier promotor bacteriano regulable. Se prefiere un promotor no filtrante.

La presente invención no se limita a un único método de diseño génico. No es necesario que el gen *osp-ipbd* se sintetice *in toto*; pueden obtenerse partes del gen de la naturaleza; puede usarse cualquier método de ingeniería genética para producir la fusión génica correcta, siempre que puedan dirigirse fácil y correctamente las mutaciones a sitios específicos en la subsecuencia de ADN de *pbd*.

Las porciones codificantes de los genes a sintetizar se diseñan a nivel de proteínas y después se codifican en ADN. Se aprovecha la ambigüedad del código genético para permitir la colocación óptima de sitios de restricción, para crear diversas distribuciones de aminoácidos y codones variegados, para minimizar el potencial de recombinación y para reducir el uso de codones que se traducen mal en la célula hospedadora.

V.F *Consideraciones Estructurales*

El diseño de la secuencia de aminoácidos para el gen *ipbd-osp* a codificar implica varias consideraciones estructurales. El diseño es en parte diferente para cada tipo de GP. En las bacterias, las OSP no son esenciales, así que no existe necesidad de que el dominio OSP de una fusión tenga ninguna de sus funciones parentales fuera de la unión en la membrana externa.

Es deseable que la OSP no limite la orientación del dominio PBD; esto no debe confundirse con la falta de limitación dentro del PBD. Cwirla *et al*. (CWIR90), Scott y Smith (SCOT90) y Devlin *et al*. (DEVL90) han mostrado que los restos variables en péptidos aleatorios presentados en los fagos deben estar libres de la influencia de la OSP del fago. En este documento se demuestra que los dominios de unión que tienen un grado de moderado a alto de limitación conformacional mostrarán una mayor especificidad y que también es posible mayor afinidad. De este modo, se recomienda tomar codones para la variegación que especifiquen aminoácidos que aparecerán en un marco bien definido. La naturaleza de los grupos laterales se modifica a través de un amplio intervalo debido a la sustitución combinatoria de múltiples aminoácidos. Las conformaciones de la cadena principal de la mayoría de los PBD de una clase dada son muy similares. El movimiento del PBD en relación con la OSP no debe restringirse. Por tanto, a menudo es apropiado incluir un enlazador flexible entre el PBD y la OSP. Dichos enlazadores flexibles pueden tomarse de proteínas de origen natural que se sabe tienen regiones flexibles. Por ejemplo, la proteína gIII de M13 contiene regiones ricas en glicinas que se piensa que permiten a los dominios amino terminales un amplio grado de libertad. Dichos enlazadores flexibles también pueden diseñarse. Los segmentos de polipéptidos que son ricos en aminoácidos GLY, ASN, SER y ASP es probable que den origen a flexibilidad. Se prefieren particularmente múltiples glicinas.

Cuando se selecciona insertar el PBD en un bucle superficial de una OSP tal como LamB, OmpA o proteína gIII de M13, existen varias consideraciones que no surgen cuando el PBD se une al extremo de una OSP. En estos casos, la OSP ejerce alguna influencia limitadora sobre el PBD; los extremos del PBD se mantienen en posiciones más o menos fijas. Puede insertarse una secuencia de ADN muy modificada en el gen osp en los codones que codifican un bucle expuesto en la superficie y seleccionar células que tengan un fenotipo de unión específica. Cuando se sintetizan las secuencias de aminoácidos identificadas (mediante cualquier medio), la limitación de la OSP se pierde y es probable que el péptido tenga una afinidad mucho menor por la diana y una especificidad mucho menor. Tan y Kaiser (TANN77) descubrieron que un modelo sintético de BPTI que contenía todos los aminoácidos de BPTI que entran en contacto con la tripsina tienen una K_d para la tripsina de $\sim 10^7$ más alta que BPTI. De este modo, se prefiere mucho que los aminoácidos modificados sean parte de un PBD en el que las limitaciones estructurales las suministre el PBD.

Se sabe que los aminoácidos adyacentes a epítopos extraños insertados en LamB influyen en las propiedades inmunológicas de estos epítopos (VAND90). Se espera que los PBD insertados en bucles de LamB, OmpA u OSP similares estén influidos por los aminoácidos del bucle y por la OSP en general. Para obtener una presentación apropiada del PBD; puede ser necesario añadir uno o más aminoácidos enlazadores entre la OSP y el PBD. Dichos enlazadores pueden tomarse de proteínas naturales o diseñarse en base a nuestro conocimiento del comportamiento estructural de los aminoácidos. Son apropiadas secuencias ricas en GLY, SER, ASN, ASP, ARG y THR. Es probable que de uno a cinco aminoácidos en cada unión otorguen el grado deseado de flexibilidad entre la OSP y el PBD.

Un sitio de inserción preferido para el gen *ipbd* en el gen *osp* del fago es uno en el que: a) el IPBD se pliegue en su forma original, b) los dominios de OSP se plieguen en sus formas originales, y c) no exista interferencia entre los dos dominios.

Si existe un modelo del fago que indique que el extremo carboxilo o amino de una OSP está expuesto al disolvente, entonces el extremo expuesto de esta OSP madura se convierte en el candidato principal para la inserción del gen *ipbd*. Es suficiente un modelo de 3D de baja resolución.

En ausencia de una estructura 3D, los extremos amino y carboxilo de la OSP madura son los mejores candidatos para la inserción del gen *ipbd*. Una fusión funcional puede necesitar restos adicionales entre los dominios de IPBD y OSP para evitar interacciones no deseadas entre estos dominios. Pueden insertarse ADN de secuencia aleatoria o ADN que codifica una secuencia específica de una proteína homóloga al IPBD o a la OSP entre el fragmento *osp* y el fragmento *ipbd* si fuera necesario.

La fusión en un dominio fronterizo con la OSP también es un buen enfoque para obtener una fusión funcional. Smith aprovechó dicha frontera cuando subclonaba ADN heterólogo en el gen *III* de f1 (SMIT85).

Los criterios para identificar dominios OSP adecuados para provocar la presentación de un IPBD son algo diferentes de los utilizados para identificar un IPBD. Cuando se identifica una OSP, el tamaño mínimo no es tan importante ya que el dominio OSP no aparecerá en la molécula de unión final ni será necesario sintetizar el gen repetidamente en cada ronda de variegación. Las principales preocupaciones en cuanto al diseño son: a) la fusión de OSP::IPBD provoca la presentación de IPBD, b) la construcción genética inicial debe ser razonablemente conveniente, y c) que el gen *osp::ipbd* sea genéticamente estable y fácilmente manipulable. Existen varios métodos para identificar dominios. Los métodos que dependen de coordinados atómicos se han revisado por Janin y Chothia (JANI85). Estos métodos usan matrices de distancias entre carbonos α (C_α), planos de división (*cf.* ROSE85) o superficie enterrada (RASH84). Chothia y colaboradores han correlacionado el comportamiento de muchas proteínas naturales con la estructura del dominio (de acuerdo con su definición). Rashin predijo correctamente la estabilidad de un dominio que comprende restos 206-316 de termolisina (VITA84, RASH84).

Muchos investigadores han usado proteolisis parcial y análisis de la secuencia de proteínas para aislar e identificar dominios estables. (Véase, por ejemplo, VITA84, POTE83, SCOT87a y PABO79). Pabo *et al*. usaron calorimetría como indicador de que el represor cI del colifago λ contiene dos dominios; después usaron proteolisis parcial para determinar la situación del dominio frontera.

Si la única información estructural disponible es la secuencia de aminoácidos de la OSP candidata, puede usarse la secuencia para predecir los giros y los bucles. Hay una alta probabilidad de que algunos de los bucles y giros se predigan correctamente (*cf.* Chou y Fasman, (CHOU74)); estas situaciones son también candidatos para la inserción del fragmento del gen *ipbd*.

En OSP bacterianas, las principales consideraciones son: a) que se presente el PBD, y b) que la proteína quimérica no sea tóxica.

A partir de modelos topológicos de OSP, puede determinarse si el extremo amino o carboxilo de la OSP está expuesto. Si esto ocurre, entonces estas son excelentes selecciones para la fusión del fragmento *osp* al fragmento *ipbd*.

El gen *lamB* se ha secuenciado y está disponible en diversos plásmidos (CLEM81, CHAR88a,b). Se han usado numerosas fusiones de fragmentos de *lamB* con diversos genes diferentes para estudiar la exportación de proteínas en *E. coli*. A partir de diversos estudios, Charbit *et al*. (CHAR88a,b) han propuesto un modelo que especifica qué restos de LamB están: a) incluidos en la membrana, b) enfrentados al periplasma, y c) enfrentados a la superficie celular; se adopta la numeración de este modelo para los aminoácidos en la proteína madura. De acuerdo con este modelo, se definen varios bucles en la superficie externa, incluyendo: 1) restos 88 hasta 111, 2) restos 145 hasta 165, y 3) 236 hasta 251.

Considerando una mini-proteína incluida en LamB. Por ejemplo, la inserción de ADN que codifica G_1NXCX_5 $XXXCX_{10}SG_{12}$ entre codones 153 y 154 de *lamB* es probable que conduzca a una gran diversidad de derivados de LamB que son expresados en la superficie de células de *E. coli*. G_1, N_2, S_{11} y G_{12} se suministran para permitir a la mini-proteína la suficiente libertad de orientación para que pueda interaccionar de forma óptima con la diana. Usando el enriquecimiento de afinidad (que implica, por ejemplo, FACS *mediante* una diana marcada de forma fluorescente, quizás a través de varias rondas de enriquecimiento), puede obtenerse una cepa (llamada, por ejemplo, BEST) que exprese un derivado de LamB particular que muestre alta afinidad por la diana predeterminada. Es probable que un octapéptido que tenga la secuencia de los restos insertados 3 hasta 10 de BEST tenga una afinidad y especificidad similares a las observadas en BEST ya que el octapéptido tiene una estructura interna que mantiene los aminoácidos en una conformación que es bastante similar en el derivado LamB y en la mini-proteína aislada.

La fusión de uno o más nuevos dominios con una proteína puede hacer la capacidad de la nueva proteína de ser exportada desde la célula diferente de la capacidad de la proteína parental. El péptido señal de la proteína de cubierta de tipo silvestre puede funcionar como polipéptido auténtico pero puede ser incapaz de dirigir la exportación de una fusión. Para utilizar la vía dependiente de Sec, puede ser necesario un péptido señal diferente. De este modo, para expresar y presentar una proteína quimérica de BPTI/gen VIII de M13, se descubrió que era necesario utilizar un péptido señal heterólogo (el de *phoA*).

Los GP que presentan péptidos que tienen alta afinidad por la diana pueden ser bastante difíciles de eluir de la diana, particularmente de una diana multivalente. (Las bacterias que están unidas muy estrechamente simplemente pueden multiplicarse *in situ*). Para los fagos, puede introducirse un sitio de escisión para una proteasa específica, tal como el Factor de coagulación sanguínea Xa, en la proteína de fusión OSP de modo que el dominio de unión pueda escindirse del paquete genético. Dicha escisión tiene la ventaja de que todos los fagos resultantes tienen OSP idénticas y por lo tanto son igualmente infectivas, incluso si el fago que presenta el polipéptido puede eluirse de la matriz de afinidad sin escisión. Esta etapa permite la recuperación de genes valiosos que de otra manera podrían perderse. Que sepamos, nadie ha descubierto o sugerido usar una proteasa específica como medio para recuperar cualquier paquete genético que convierta información o para convertir una población de fagos que tengan infectividad variable en fagos que tengan infectividad idéntica.

IV.G. *Síntesis de Insertos de Genes*

La presente invención no se limita a ningún método o estrategia particular de síntesis o construcción de ADN. Pueden usarse sintetizadores de ADN convencionales, con las modificaciones de reactivos apropiadas para la producción de ADN variegado (similar al usado actualmente para la producción de sondas mixtas).

Los *genes osp-pbd* pueden crearse insertando ADNvg en un gen parental existente, tal como el *osp-ipbd* que haya demostrado ser presentable por un GP transformado de forma adecuada. La presente invención no se limita a ningún método de introducción del ADNvg particular, por ejemplo mutagénesis de cassette o mutagénesis dirigida por oligonucleótidos de cadena sencilla.

IV.H. *Vector de Clonación Operativo*

El vector de clonación operativo (OCV) es un ácido nucleico replicable usado para introducir el gen *ipbd-osp* o *ipbd-osp* quimérico en el paquete genético. Cuando el paquete genético es un virus, éste puede servir como su propio OCV. Para las células y esporas, el OCV puede ser un plásmido, un virus, un fagémido o un cromosoma.

IV.I. *Transformación de células*

Cuando el GP es una célula, la población de GP se crea transformando las células con OCV adecuados. Cuando el GP es un fago, el fago se manipula genéticamente y después se transfecta en células hospedadoras adecuadas para su amplificación. Cuando el GP es una espora, se transforman células capaces de esporulación con el OCV mientras está en un estado metabólico normal, y después se induce la esporulación para provocar que se presente el OSP-PBD. La presente invención no se limita a ningún método de transformación de células con ADN.

Las células transformadas se cultivan en primer lugar en condiciones no selectivas que permiten la expresión de genes de plásmidos y después se seleccionan para destruir células no transformadas. Las células transformadas se inducen después para que expresen el gen *osp-pbd* en el nivel de inducción apropiado. Después, se recogen los GP que tienen el IPBD o PBD mediante métodos apropiados para el GP a mano, generalmente centrifugado para sedimentar GP y resuspensión de los sedimentos en medio estéril (células) o tampón (esporas o fago). Después están preparadas para la verificación de que la estrategia de presentación era un éxito (donde todos los GP presentan un IPBD "de ensayo") o para la selección por afinidad (donde los GP presentan diversos PBD diferentes).

IV.J. *Verificación de la Estrategia de Presentación:*

Se ensayan los paquetes recogidos para determinar si el IPBD está presente en la superficie. En cualquier ensayo de GP para la presencia de IPBD en la superficie del GP, se incluye cualquier ión o cofactor que se sabe son esenciales para la estabilidad de IPBD o AfM(IPBD) a los niveles apropiados. Los ensayos pueden realizarse, por ejemplo, mediante a) marcado de afinidad, b) enzimáticamente, c) espectrofotométricamente, d) mediante separación por afinidad, o e) mediante precipitación por afinidad. En esta etapa el AfM(IPBD) es uno tomado para tener una afinidad fuerte (preferiblemente, $K_d < 10^{-11}$ M) por la molécula de IPBD y poca o ninguna afinidad por el wtGP.

V. *Selección por afinidad de mutantes de unión a la diana*

V.A. *Tecnología de Separación por Afinidad. En Líneas Generales*

La separación por afinidad se usa inicialmente en la presente invención para verificar que el sistema de presentación funciona, *es decir*, que una proteína de superficie externa quimérica se ha expresado y transportado a la superficie del paquete genético y está orientada de modo que el dominio de unión insertado sea accesible al material diana. Cuando se usa para este fin, el dominio de unión es un domino de unión conocido para una diana particular y esta diana es la molécula de afinidad usada en el proceso de separación por afinidad. Por ejemplo, un sistema de presentación puede validarse usando ADN de inserción que codifica BPTI en un gen que codifica una proteína de la superficie externa del paquete genético de interés, y ensayando la unión a anhidrotripsina, que normalmente se une mediante BPTI.

Si los paquetes genéticos se unen a la diana, entonces tenemos la confirmación de que el dominio de unión correspondiente se presenta de hecho en el paquete genético. Los paquetes que presentan el dominio de unión (y por lo tanto se unen a la diana) se separan de los que no lo hacen.

Una vez que el sistema de presentación se valida, es posible usar una población variegada de paquetes genéticos que presentan diversos dominios de unión potencial diferentes, y usar la tecnología de separación por afinidad para determinar en qué grado se unen a una o más dianas. Esta diana no necesita ser una unida mediante un dominio de unión conocido que sea parental para los dominios de unión presentados, *es decir*, puede seleccionarse la unión a una nueva diana.

La expresión "medios de separación por afinidad" incluye, aunque sin limitación: a) cromatografía en columna de afinidad, b) elución en lotes a partir de un material de matriz de afinidad, c) elución en lotes a partir de un material de afinidad unido a una placa, d) selección celular activada por fluorescencia, y e) electroforesis en presencia del material diana. "Material de afinidad" se usa para referirse a un material con afinidad por el material a purificar, llamado el "analito". En la mayoría de los casos, la asociación del material de afinidad y el analito es reversible de modo que el analito puede liberarse del material de afinidad una vez que las impurezas se han retirado mediante lavado.

V.B. *Cromatografía de Afinidad. En Líneas Generales*

La cromatografía en columna de afinidad, la elución en lotes a partir de un material de matriz de afinidad contenido en algún recipiente y la elución en lotes de una placa son muy similares y se denominarán en lo sucesivo en este documento "cromatografía de afinidad".

Si se va a usar cromatografía de afinidad, entonces:

1) las moléculas del material diana deben ser del tamaño y la reactividad química suficientes para aplicarse a un soporte sólido adecuado para la separación por afinidad,

2) después de la aplicación a una matriz, el material diana preferiblemente no reacciona con el agua,

3) después de la aplicación a una matriz, el material diana preferiblemente no se une o degrada proteínas de manera no específica, y

4) las moléculas del material diana deben ser lo suficientemente grandes para que la unión del material a una matriz permita la suficiente área superficial inalterada (generalmente de al menos 500 $Å^2$, excluyendo el átomo que está conectado al enlazador) para la unión de proteínas.

La cromatografía de afinidad es el medio de separación preferido, pero pueden usarse también FACS, electroforesis u otros medios.

La presente invención hace uso de la separación por afinidad de células bacterianas o virus bacterianos (u otros paquetes genéticos) para enriquecer una población en aquellas células o virus que tienen genes que codifican proteínas con propiedades de unión deseables.

V.C. *Materiales diana*

La presente invención puede usarse para seleccionar dominios de unión que se unen a uno o más materiales diana y/o que no consiguen unirse a uno o más materiales diana. Por supuesto, la especificidad es la capacidad de una molécula de unión para unirse fuertemente a un conjunto limitado de materiales diana, mientras se une más débilmente o no se une en absoluto a otro conjunto de materiales diana de los que debe distinguirse el primer conjunto.

Los materiales diana pueden ser macromoléculas orgánicas, tales como polipéptidos, lípidos, ácidos polinucleicos y polisacáridos, pero no se limitan a estos. La presente invención no se limita sin embargo a ninguno de los materiales diana identificados anteriormente. La única limitación es que el material diana sea adecuado para la separación por afinidad. De este modo, cualquier molécula que sea estable en el disolvente acuoso puede usarse como diana.

Las serinas proteasas tales como la elastasa neutrófila humana (HNE) son una clase especialmente interesante de materiales diana potenciales. La serina proteasas se encuentran en todos los organismos vivos y desempeñan papeles vitales en procesos tales como digestión, coagulación sanguínea, fibrinolisis, respuesta inmune, fertilización y procesamiento post-traduccional de hormonas peptídicas. Aunque el papel que desempeñan estas enzimas es vital, la actividad proteolítica incontrolada o inapropiada puede ser muy perjudicial.

V.D. *Inmovilización o Marcado del Material Diana*

Para cromatografía, FACS o electroforesis puede existir una necesidad de unir covalentemente el material diana a una segunda entidad química. Para cromatografía, la segunda entidad es una matriz, para FACS la segunda entidad es un tinte fluorescente y para electroforesis la segunda entidad es una molécula fuertemente cargada. En muchos casos, no se requiere acoplamiento ya que el material diana ya tiene la propiedad deseada de: a) inmovilidad, b) fluorescencia o c) carga. En otros casos, se requiere el acoplamiento químico o físico.

No es necesario que el auténtico material diana se use para preparar el análogo inmovilizado o marcado que se usará en la separación por afinidad; en su lugar, pueden ser más convenientes análogos reactivos adecuados del material

diana. Los materiales diana que no tienen grupos funcionales reactivos pueden inmovilizarse creando en primer lugar un grupo funcional reactivo a través del uso de algún reactivo poderoso, tal como un halógeno. En algunos casos, los grupos reactivos del material diana auténtico pueden ocupar una parte de la molécula diana que debería dejarse inalterada. En estos casos, pueden introducirse grupos funcionales adicionales mediante química de síntesis.

Dos métodos muy generales de inmovilización se usan ampliamente. El primero es biotinilar el compuesto de interés y después unir el derivado biotinilado a avidina inmovilizada. El segundo método es generar anticuerpos para el material diana, inmovilizar los anticuerpos mediante cualquiera de muchos métodos y después unir el material diana a los anticuerpos inmovilizados. El uso de anticuerpos es más adecuado para materiales diana más grandes; los anticuerpos pueden absorber completamente las dianas pequeñas (las que comprenden, por ejemplo, 10 o menos átomos no de hidrógeno) de modo que queda muy poca diana expuesta en el complejo diana-anticuerpo.

También puede usarse la inmovilización de moléculas hidrófobas sin recurrir a anticuerpos. Un compuesto, tal como 2,3,3-trimetildecano se mezcla con una matriz precursora, tal como alginato de sodio y la mezcla se extruye a una solución de endurecimiento. Las perlas resultantes tendrán 2,3,3-trimetildecano dispersado en ellas y expuesto en la superficie.

Otros métodos de inmovilización dependen de la presencia de funcionalidades químicas particulares. Un polipéptido presentará -NH_2 (N-terminal; Lisinas), -COOH (C-terminal; Ácidos Aspártico y glutámico), -OH (Serinas; Trioninas; Tirosinas) y -SH (Cisteinas). Para la reactividad de cadenas laterales de aminoácidos véase CREI84. Un polisacárido tiene grupos -OH libres, como el ADN, que tiene una estructura principal de azúcares.

Las matrices adecuadas para su uso como materiales de soporte incluyen poliestireno, vidrio, agarosa y otros soportes cromatográficos y pueden fabricarse en forma de perlas, láminas, columnas, pocillos y otras formas según se desee.

De forma temprana en el proceso de selección, pueden aplicarse concentraciones relativamente altas de materiales diana a la matriz para facilitar la unión; las concentraciones de la diana pueden reducirse posteriormente para seleccionar SBD de alta afinidad.

V.E. *Elución de PaquetesGenéticos que llevan PBD de baja Afinidad*

La población de GP se aplica a una matriz de afinidad en condiciones compatibles con el uso pretendido de que la proteína de unión y la población se fraccionen mediante el paso de un gradiente de algún soluto por la columna. El proceso enriquece los PBD que tienen afinidad por la diana y para los que la afinidad por la diana está menos afectada por los eluyentes usados. Las fracciones enriquecidas son las que contienen GP viables que se eluyen de las columnas a una mayor concentración del eluyente.

Los eluyentes preferiblemente son capaces de debilitar las interacciones no covalentes entre los PBD presentados y el material diana inmovilizado. Preferiblemente, los eluyentes no destruyen el paquete genético; el mensaje genético correspondiente a las mini-proteínas de éxito se amplifica más convenientemente reproduciendo el paquete genético, en lugar de mediante procedimientos *in vitro* tales como PCR. La lista de eluyentes potenciales incluyen sales (incluyendo Na+, NH_4+, Rb+, SO_4-, H2PO_4-, citrato, K+, Li+, Cs+, HSO_4-, CO_3-, Ca++, Sr++, Cl-, PO_4-, HCO_3-, Mg++, Ba++, Br-, HPO_4- y acetato), ácido, calor, compuestos que se sabe se unen a la diana y material diana soluble (o análogo de los mismos).

Los paquetes genéticos no eluidos contienen ADN que codifica dominios de unión que tienen una afinidad lo suficientemente alta por el material diana para resistir las condiciones de elución. El ADN que codifica tales dominios de unión de éxito puede recuperarse de diversas maneras. Preferiblemente, los paquetes genéticos unidos se eluyen simplemente por medio de un cambio en las condiciones de elución. Como alternativa, puede cultivarse el paquete genético *in situ*, o extraerse la matriz que contiene la diana con fenol (u otro disolvente adecuado) y amplificar el ADN mediante PCR o mediante técnicas de ADN recombinante. Además, si se ha manipulado genéticamente un sitio para una proteasa específica en el vector de presentación, se usa la proteasa específica para escindir el dominio de unión del GP.

La unión no específica de la matriz, etc. puede simplificarse o reducirse mediante técnicas bien conocidas en la técnica de separación por afinidad.

V.F. *Recuperación de los paquetes*

La recuperación de los paquetes que presentan unión a una columna de afinidad puede conseguirse de varias maneras, incluyendo:

1) recoger fracciones eluidas de la columna con un gradiente como se ha descrito anteriormente; las fracciones que se eluyen después en el gradiente contienen GP más enriquecidos en genes que codifican PBD con alta afinidad por la columna,

2) eluir la columna con el material diana en forma soluble,

3) inundar la matriz con un medio nutritivo y cultivar los paquetes deseados *in situ*,

4) retirar partes de la matriz y usarlas para inocular medio de cultivo,

5) degradar química o enzimáticamente el enlace que mantiene la diana en la matriz de modo que los GP que siguen unidos a la diana se eluyan, o

6) degradar los paquetes y recuperar el ADN con fenol u otro disolvente adecuado; el ADN recuperado se usa para transformar células que regeneran GP.

Es posible utilizar combinaciones de estos métodos. Debe recordarse que lo que queremos recuperar de la matriz de afinidad no es el GP *per se*, sino la información en su interior. La recuperación de GP viables es muy preferente, pero la recuperación del material genético es esencial. Si las células, esporas o virones se unen de forma irreversible a la matriz pero no se destruyen, puede recuperarse la información a través de una división celular, germinación o infección *in situ*, respectivamente. La degradación proteolítica de los paquetes y la recuperación del ADN no se prefiere.

V.G. *Amplificación de los Paquetes Enriquecidos*

Los GP viables que tienen el rasgo de unión seleccionado se amplifican mediante un cultivo en un medio adecuado, o en el caso del fago, por infección en un hospedador cultivado de este modo. Si los GP se han inactivado mediante cromatografía, el OCV que lleva el gen *osp-pbb* se recupera del GP y se introduce en un nuevo hospedador viable.

V.H. *Caracterización del supuesto SBD*

Para uno o más aislados clonales, se subclona el fragmento del gen *sbd*, sin el fragmento *osp* en un vector de expresión tal que cada SBD puede producirse en forma de proteína libre. Pueden realizarse mediciones físicas de la fuerza de unión para cada proteína SBD libre mediante cualquier método adecuado.

Si se descubre que la unión aún no es suficiente, se decide que restos del SBD (ahora un nuevo PPBD) modificar a continuación. Si la unión es suficiente, entonces se tiene ahora un vector de expresión que tiene un gen que codifica la nueva proteína de unión deseada.

V.I. *Selecciones de Unión*

Puede modificarse la separación por afinidad del método descrito para seleccionar la molécula que se une al material A pero no al material B, o que se une a A y a B alternativa o simultáneamente.

V.J. *Manipulación de Antagonistas*

Puede ser deseable proporcionar un antagonista de una enzima o receptor. Esto puede conseguirse preparando una molécula que impide al sustrato o agonista natural alcanzar el sitio activo. Las moléculas que se unen directamente al sitio activo pueden ser agonistas o antagonistas. De este modo se adopta la siguiente estrategia. Se considera a las enzimas y a los receptores conjuntamente con la designación TER (Enzima o Receptor Diana).

Para la mayoría de los TER, existen inhibidores químicos que bloquean el sitio activo. Normalmente, estos compuestos químicos son útiles solamente como herramientas de investigación debido a su alta toxicidad. Se preparan dos matrices de afinidad: una con el TER activo y otra con el TER bloqueado. Se prepara una población variegada de GP (PBD) y se seleccionan SBP que se unen a las dos formas de la enzima, obteniendo SDP que no se unen al sitio activo. Se espera descubrir SBD que se unan en diferentes sitios de la superficie de la enzima. Se fusionan pares de los genes *sbd* con un segmento peptídico entre ambos. Por ejemplo, si SBD-1 y SBD-2 son dominios de unión que muestran afinidad alta por la enzima diana y para los cuales la unión es no competitiva, entonces el gen *sbd-1::enlazador::sbd-2* codifica una proteína de dos dominios que mostrará alta afinidad por la diana. Se realizan varias fusiones que tienen diversos SBD y diversos enlazadores. Dichos compuestos pueden una probabilidad razonable de ser un antagonista de la enzima diana.

VI. *Aprovechamiento de los dominios de unión de éxito y de los ADN correspondientes*

Aunque el SBD puede producirse mediante técnicas de ADN recombinante, una ventaja inherente del uso de una mini-proteína como IPBD es que es probable que el SBD también se comporte como una proteína y sea obtenible por medio de una síntesis química. (La expresión "síntesis química", como se usa en este documento incluye el uso de agentes enzimáticos en un entorno libre de células).

También debe entenderse que las mini-proteínas obtenidas mediante el método de la presente invención pueden tomarse como compuestos de cabecera para una serie de homólogos que contienen aminoácidos de origen no natural y grupos diferentes de aminoácidos. Por ejemplo, puede sintetizarse una serie de homólogos en los cuales cada miembro de la serie tiene un aminoácido sustituido por su enantiómero D. También pueden prepararse homólogos que contienen

constituyentes tales como β alanina, ácido aminobutírico, 3-hidroxiprolina, ácido 2-aminoadípico, N-etilasperagina, norvalina, *etc*.; Estos podían ensayarse para la unión y otras propiedades de interés, tales como estabilidad y toxicidad.

Los péptidos pueden sintetizarse químicamente en solución o sobre soportes. Pueden emplearse diversas combinaciones de síntesis por etapas y condensación de fragmentos.

Durante la síntesis las cadenas laterales de aminoácidos se protegen para evitar la ramificación. Varios grupos protectores diferentes son útiles para la protección de los grupos tiol de las cisteínas:

1) 4-metoxibenzil (MBzl; Mob)(NIS82; ZAFA88), eliminable con HF;

2) acetamidometil (Acm) (NIS82; NIS86; BECK89c), eliminable con yodo; iones de mercurio (*por ejemplo*, acetato mercúrico); nitrato de plata; y

3) S-para-metoxibenzilo (HOUG84).

Otros grupos protectores de tiol pueden encontrarse en trabajos de referencia convencionales tales como Greene, PROTECTIVE GROUPS IN ORGANIC SYNTHESIS (1981).

Una vez que la cadena polipeptídica se ha sintetizado, pueden formarse puentes disulfuro. Los agentes oxidantes posibles incluyen el aire (HOUG84; NISH86), ferricianuro (NISH82; HOUG84), yodo (NISH82) y ácido perfórmico (HOUG84). La temperatura, el pH, el disolvente y productos químicos caotrópicos pueden afectar al curso de la oxidación.

Se han sintetizado gran cantidad de micro-proteínas con una pluralidad de puentes disulfuro en forma biológicamente activa: conotoxina G1 (13AA, 4 Cys)(NISH-82); enterotoxina ST estable al calor (18AA, 6 Cys) (HOUG84); análogos de ST (BHAT86); Ω-conotoxina GVIA (27AA, 6Cys) (N-ISH86; RIVI87b); Ω-conotoxina MVIIA (27 AA, 6 Cys) (OLIV87b); α-conotoxina SI (13 AA, 4 Cys) (ZAFAB8); μ-conotoxina IIIa (22AA, 6 Cys) (BECK89c, CRUZ89, HATA90). Algunas veces, el polipéptido se pliega de forma natural de modo que se forman los puentes disulfuro correctos. Otras veces, debe ayudársele mediante el uso de diferentes grupos protectores que pueden retirarse para cada par de cisteínas.

Los dominios de unión de éxito pueden usarse, en solitario o como parte de una proteína más grande, para cualquier fin para el que sean adecuadas proteínas de unión, incluyendo aislamiento, detección de materiales diana. Para promover este fin, las nuevas proteínas de unión pueden acoplarse directa o indirectamente, covalente o no covalentemente a una marca, vehículo o soporte.

Cuando se usan como producto farmacéutico, las nuevas proteínas de unión pueden estar encerradas junto con vehículos o adyuvantes apropiados.

Ejemplo comparativo I

Diseño y mutagénesis de una micro-proteína de clase 1

Para obtener una biblioteca de dominios de unión que estén limitados conformacionalmente por un único puente disulfuro, se inserta ADN que codifica la siguiente familia de micro-proteínas en el gen que codifica una OSP adecuada.

$$X_1\text{-}C\text{-}X_2\text{-}X_3\text{-}X_4\text{-}X_5\text{-}C\text{-}X_6$$

Donde ⌊——⌋ indica puente disulfuro. Los puentes disulfuro no se forman normalmente entre cisteínas que son consecutivas en la cadena polipeptídica. Uno o más de los restos indicados anteriormente como X_n se modificarán ampliamente para obtener nueva unión. Puede haber uno o más aminoácidos que precedan a X_1 o que sigan a X_6, sin embargo, los restos antes de X_1 o después de X_6 no estarán limitados significativamente por el puente disulfuro programado y es menos ventajoso modificar estos restos lejanos no unidos por puentes. El último resto X está conectado con la OSP del paquete genético.

X_1, X_2, X_3, X_4, X_5, y X_6 pueden modificarse de forma independiente; *es decir*, puede usarse un esquema diferente de variegación en cada posición. X_1 y X_6 son los restos menos limitados y pueden modificarse menos que las otras posiciones.

X_1 y X_6 pueden ser, por ejemplo, uno de los aminoácidos [E, K, T y A]; este conjunto de aminoácidos se prefiere porque: a) se proporciona la posibilidad de aminoácidos cargados positivamente, cargados negativamente y neutros; b) estos aminoácidos pueden proporcionarse en una proporción 1:1:1:1 mediante el codón RMG (R = A y G equimolares, M = A y C equimolares) esto aminoácidos permiten un procesamiento adecuado mediante peptidasas de señal.

En una realización preferida, X_2, X_3, X_4 y X_5 se variegan inicialmente codificando cada uno mediante el codón NNT, que codifica el conjunto de sustitución [F, S, Y, C, L, P, H, R, I, T, N, V, A, D y G].

Las ventajas del NNT sobre el codón NNK se hacen cada vez más evidentes a medida que la cantidad de codones variegados aumenta. Las Tablas 10 y 130 comparan bibliotecas en las cuales se han modificado 6 codones mediante los codones NNT o NNK. NNT codifica 15 aminoácidos diferentes y solamente 16 secuencias de ADN. Por lo tanto, existen $1{,}139 \cdot 10^7$ secuencias de aminoácidos que no son interrupciones y solamente $1{,}678 \cdot 10^7$ secuencias de ADN. Una biblioteca de 10^8 transformantes independientes contendrá el 99% de todas las secuencias posibles. La biblioteca NNK contiene $6{,}4 \cdot 16^7$ secuencias, pero el muestreo completo requiere una cantidad mucho mayor de transformantes independientes.

Esta secuencia puede presentarse en forma de fusión a la proteína del gen III de M13 usando el promotor del gen III de M13 y la secuencia señal nativos. La secuencia de la proteína del gen III de M13, desde el resto 16 al 23 es $S_{16}HSAETVE_{23}$; la peptidasa de señal I escinde a partir del S_{18}. Este segmento se sustituye con

$$S_{16}GA_{18}AEGX_1CX_2X_3X_4X_5X_6SYIEGRVIETVE.$$

Obsérvese que el cambio de $H_{17}S_{18}$ por GA no altera la infectividad del fago. Esto es útil para insertar un sitio de reconocimiento de F.Xa/sitio de escisión bovino (YIEGR/VI) entre el PBD y la proteína madura III; esto no solamente permite la libertad de orientación para el PBD sino que también permite la escisión del PBD del GP.

Una biblioteca de fagos, en la que NNG codifica X_1, X_2, X_3 y X_6 (permitiendo F, S, Y, C, L, P, H, R, V, T, N, V, A, D y G) y en la NNG codifica X_3 y X_4 (permitiendo L, S, W, P, Q, R, M, T, K, V, A, E y G) se denomina TN2. Esta biblioteca presenta aproximadamente 8,55 x 10^6 micro-proteínas codificadas por aproximadamente 1,5 x 10^7 secuencias de ADN. Se usa NNG en la tercera y cuarta posiciones variables (las posiciones centrales del bucle cerrado con puentes disulfuro) al menos en parte para evitar la posibilidad de la aparición de cisteínas en esas posiciones.

Devlin, *et al*., seleccionaron 10^7 transformantes, cada uno de los cuales podía presentar uno de 10^{12} pentadecapéptidos aleatorios, para la afinidad con estreptavidina y descubrieron 20 aislados de fago de unión a estreptavidina, con ocho secuencias únicas ("A"-"I"). Todos contenían HP; 15/20 HPQ; y 6/20 HPQF, aunque en diferentes posiciones dentro del pentadecapéptido. Los aislados que se encontraban con más frecuencia eran D(5), I(4) y A(3), que carecían completamente de cisteína. Sin embargo, dos aislados posibles "E"(1) y "F"(2), incluían un par de cisteínas colocadas de modo que era posible la formación de un puente disulfuro. La secuencia de estos aislados se proporciona en la Tabla 820.

Se descubrió que nuestra librería TN2 debe incluir una supuesta micro-proteína, HPQ, lo suficientemente similar a los péptidos "E" y "F" de Devlin para que tenga el potencial de mostrar actividad de unión a estreptavidina. HPQ comprende la secuencia aminoterminal AEG común a todos los miembros de la biblioteca TN2, seguida de la secuencia PCHPQFCQ que tiene el potencial de formar un puente disulfuro en un tramo de cuatro, seguida de una serina (S) y un sitio de reconocimiento del factor Xa bovino (YIEGR/IV) (véase Tabla 820). Los experimentos piloto mostraron que la unión del fago que tenía HPQ a estreptavidina era comparable a la del aislado "F" de Devlin; ambos estaban muy poco por encima del trasfondo (1,7x). Por lo tanto, se exploró la biblioteca TN2 frente a estreptavidina inmovilizada.

La estreptavidina está disponible en forma de proteína libre (Pierce) con una actividad específica de 14,6 unidades por mg (1 unidad se unirá a 1 μg de biotina). Se preparó una solución madre de 1 mg por ml en PBS que contenía azida al 0,01%. Se añaden 100 μl de solución madre StrAv a cada pocillo de 250 μl de capacidad de placas Inmulon (Nº 4) y se incuba durante una noche a 4ºC. La solución madre se retira y se sustituye con 250 μl de PBS que contiene BSA a una concentración de 1 mg/ml y se deja a 4ºC durante una hora adicional. Antes de su uso en un ensayo de unión de fagos, los pocillos se lavan rápidamente cinco veces con 250 μl de PBS que contenía Tween al 0,1%.

A cada pocillo recubierto con StrAv se le añaden 100 μl de tampón de unión (PBS con 1 mg por ml de BSA) que contenía una cantidad conocida de fago (10^{11} pfu de la biblioteca TN2). La incubación continúa durante 1 hora a temperatura ambiente seguida de la retirada del fago no unido y 10 lavados rápidos con PBS con Tween al 0,1% y después lavados adicionales con tampones citrato de pH 7, 6 y 5 para retirar la unión no específica. El fago unido se eluye con 250 μl de tampón citrato de pH 2 que contenía 1 mg por ml de BSA y neutralización con 60 μl de tris 1 M a pH 8. El eluato se usó para infectar células bacterianas que generaban una nueva solución madre de fago que puede usarse para una ronda adicional de unión, lavado y elución. Los ciclos de potenciación se repitieron dos veces más (tres en total) después de los cuales se secuenciaron una cantidad de fagos individuales y se ensayaron como aislados clonales. La cantidad de fagos presentes en cada etapa se determina como unidades formadoras de placas (pfu) después de las diluciones apropiadas y siembra en placas en un cultivo de F' que contenía *E. coli*.

La tabla 838 muestra las secuencias peptídicas que se descubrió que se unían a StrAv y su frecuencia en las tomas aleatorias realizadas a partir de la reserva de fagos final (ronda 3).

El segmento entre cisteínas de todas las supuestas micro-proteínas examinadas contenía el motivo HPQF. El resto variable antes de la primera cisteína podía contener cualquiera de {F,S,Y,C,L,P,H,R,I,T,N,V,A,D,G}; los restos seleccionados fueron {Y,H,L,D,N} aunque el fago HPQ tiene P. El resto variable después de la segunda cisteína también

podía haber tenido {F,S,Y,C,L,P,H,R,I,T,N,V,A,D,G}; los restos seleccionados fueron {P,S,G,R,V} aunque el fago HPQ tiene Q. La relativamente mala unión del fago HPQ podía deberse a P_4 o a Q_{13}, o a ambos.

En un experimento de control, la biblioteca TN2 se exploró de igual manera a la que se ha mostrado anteriormente, pero siendo la proteína diana el agente bloqueante BSA. Después de tres rondas de unión, elución y amplificación, se tomaron diez y seis placas de fago aleatorias y se secuenciaron. La mitad de los clones demostraron carencia del inserto (8/16), la otra mitad tenían las secuencias que se muestran en la tabla 839. No hay consenso para esta colección.

Se ha presentado una micro-proteína relacionada, HPQ6, en el fago. Es idéntica a HPQ excepto por la sustitución de CHPQFC con CHPOFPRC (véase la Tabla 820). Cuando se presenta, HPQ6 tiene una afinidad sustancialmente mayor por la estreptavidina que HPQ o el aislado "F" de Devlin. (El aislado "E" de Devlin no se estudió). El tratamiento con ditiotreitol (DTT) redujo drásticamente la unión del fago HPQ6 (pero no del fago de control) a estreptavidina, sugiriendo que la presencia de un puente disulfuro dentro del péptido presentado era necesaria para una buena unión. En vista de los resultados de la exploración de la biblioteca TN2, es probable que la unión del fago HPQ6 pudiera mejorarse adicionalmente cambiando P_4 por uno de {Y,H,L,D,N} y/o cambiando Q_{13} por uno de {P,S,G,R,V}.

Ejemplo comparativo II

Una unidad CYS::hélice::giro::cadena::CYS

La micro-proteína de Clase 2 parental puede ser una micro-proteína de Clase 2 de origen natural. También puede ser un dominio de una proteína mayor cuya estructura satisface o puede modificarse para satisfacer los criterios de una micro-proteína de clase 2. La modificación puede ser una modificación sencilla, tal como la introducción de una cisteína (o un par de cisteínas) en la base de una estructura de horquilla, de modo que la horquilla pueda cerrarse con un puente disulfuro, o una modificación más elaborada, como la modificación de restos intermedios para conseguir la estructura de horquilla. La micro-proteína de clase 2 parental también puede ser una composición de estructuras de dos o más proteínas de origen natural, por ejemplo, una α hélice de una proteína y una lámina β de una segunda proteína.

Un motivo de micro-proteína de uso potencial comprende un bucle disulfuro que encierra a una hélice, un giro y una cadena de retorno. Dicha estructura puede diseñarse o puede obtenerse a partir de una proteína de estructura 3D conocida. La neurotoxina del escorpión, variante 3 (ALMA83a, ALMA83b) (en lo sucesivo en este documento ScorpTx) contiene una estructura que se representa en la Figura 1, que comprende una hélice (restos N22 hasta N33), un giro (restos 33 hasta 35) y una cadena de retorno (restos 36 hasta 41). La ScorpTx contiene puentes disulfuro que unen los restos 12-65, 16-41, 25-46 y 29-48. CYS_{25} y CYS_{41} están bastante cerca y pueden unirse mediante un puente disulfuro sin desorganizar la cadena principal. La Figura 1 muestra CYS_{25} unido a CYS_{41}. Además, el CYS_{29} se ha cambiado por GLN. Se espera que se forme un puente disulfuro entre 25 y 41 y que se forme la hélice que se muestra. Se sabe que la secuencia de aminoácidos que se muestra es altamente compatible con esta estructura. La presencia de GLY_{35}, GLY_{36} y $GLKY_{39}$ proporciona al giro y a la lámina extendida la suficiente flexibilidad para alojar cualquier cambio necesario alrededor de CYS_{41} para formar el puente disulfuro.

A partir del examen de esta estructura (como se descubre en la entrada 1SN3 del Brookhaven Protein Data Bank), se observa que serían preferibles los siguientes conjuntos de restos para la variegación.

CONJUNTO 1

Resto	Codón	Aminoácidos permitidos	Naa/Nadn
1) T_{27}	NNG	L^2R^2MVSPTAQKEWG.	13/15
2) E_{28}	VHG	LMVPTAGKE	9/9
3) A_{31}	VHG	LMVPTAGKE	9/9
4) K_{32}	VHG	LMVPTAGKE	9/9
5) G_{24}	NNG	L^2R^2MVSPTAQKEWG.	13/15
6) E_{23}	VHG	LMVPTAGKE	9/9
7) Q_{34}	VAS	HQNKED	6/6

Nota: Los exponentes sobre los aminoácidos indican la multiplicididad de los codones.

Las posiciones 27, 28, 31, 32, 24 y 23 comprenden una cara de la hélice. En cada uno de esos puntos hemos tomado un codón de variegación que a) incluye el aminoácido parental, b) incluye un conjunto de restos que tienen una predominancia de restos que favorecen la formación de la hélice, c) proporciona una amplia variedad de aminoácidos

y d) conduce a una distribución lo más plana posible. La posición 34 es parte de un giro. El grupo lateral del resto 34 podía interaccionar con moléculas que entran en contacto con los grupos laterales de los restos 27, 28, 31, 32, 24 y 23. De este modo, se permite la variegación en ese punto y se proporcionan aminoácidos que son compatibles con los giros. La variegación que se muestra conduce a $6{,}65 \cdot 10^6$ secuencias de aminoácidos codificadas por $8{,}85 \cdot 10^6$ secuencias de ADN.

CONJUNTO 2

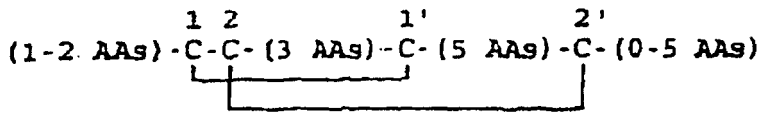
Resto	Codón	Aminoácidos permitidos	Naa/Nadn
1) D_{26}	VHS	L^2|$MV^2P^2T^2A^2$HQNKDE	13/18
2) T_{27}	NNG	L^2R^2MVSPTAQKEWG.	13/15
[3]) K_{30}	VHG	KEQPTALMV	9/9
4) A_{31}	VHG	KEQPTALMV	9/9
5) K_{32}	VHG	LMVPTAGKE	9/9
6) S_{37}	RRT	SNDG	4/4
7) Y_{38}	NHT	YSFHPLNTIDAV	9/9

Las posiciones 26, 27, 30, 31 y 32 se variegan para potenciar los aminoácidos que favorecen la formación de la hélice en la población. Los restos 37 y 38 están en la cadena de retorno de modo que se toman diferentes codones de variegación. Esta variegación permite $4{,}43 \cdot 10^6$ secuencias de aminoácidos y $7{,}08 \cdot 10^6$ secuencias de ADN. Por lo tanto, una biblioteca que sea una realización de este esquema puede muestrearse de forma muy eficaz.

Ejemplo III

Diseño y mutagénesis de micro-proteína de clase 3. Micro-proteínas Parentales con Dos Puentes Disulfuro

Pueden modelarse micro-proteínas con dos puentes disulfuro a partir de α-conotoxinas, *por ejemplo*, GI, GIA, GII, MI y SI. Éstas tienen la siguiente estructura conservada:



Hashimoto *et al*. (HASH85) presentaron la síntesis de 24 análogos de α-conotoxinas GI, GII y MI. Usando el esquema de numeración para GI (CYS en las posiciones 2, 3, 7 y 13), Hashimoto *et al*. presentaron alteraciones en 4, 8, 10 y 12 que permiten que la proteína sea tóxica. Almquist *et al*. (ALMQ89) sintetizaron α-conotoxina GI [des-GLU_1] y veinte análogos. Descubrieron que sustituyendo GLY por PRO_5 se daba origen a dos isómeros, quizás relacionados con la formación de diferentes puentes disulfuro. Descubrieron varias sustituciones en los restos 8 a 11 que permitían que la proteína fuera tóxica. Zafaralla *et al*. (ZAFA88) descubrieron que la sustitución de PRO en la posición 9 da una proteína activa. Cada uno de los grupos mencionados se usaba solamente en toxicidad *in vivo* como un ensayo para la actividad. A partir de dichos estudios, puede inferirse que una proteína activa tiene la estructura 3D parental, pero no puede inferirse que una proteína inactiva carezca de esta estructura 3D parental.

Pardi *et al*. (PARD89) determinaron la estructura 3D de una α-conotoxina GI obtenida a partir de veneno mediante RMN. Kobayashi *et al*. (KOBA89) han presentado una estructura 3D de la α-conotoxina GI a partir de datos de RMN que concuerda con la de PARD89. Se refiere a la Figura 5 de Pardi *et al*.

Se sabe que el resto GLU_1 se adapta a GLU, ARG, e ILE en análogos u homólogos conocidos. Un codón de variegación preferido es NNG que proporciona el conjunto de aminoácidos [L^2R^2MVSPTA-QKEWG<detención>]. A partir de la Figura 5 de Pardi *et al*. se observa que el grupo lateral de GLU_1 se proyecta en la misma región que la cadena que comprende los restos 9 a 12. Los restos 2 y 3 son cisteínas y no deben modificarse. El grupo lateral del resto 4 está lejos del resto 9 a 12; por lo tanto, se prefiere no modificar este resto hasta una ronda posterior. PRO_5, puede ser necesaria para provocar que se formen los puentes disulfuro correctos; cuando GLY sustituyó en este punto, el péptido se plegó de dos formas, ninguna de las cuales es tóxica. Se permite modificar PRO_5, pero se prefiere no hacerlo en la primera ronda.

No se han presentado sustituciones en ALA_6. Un codón de variegación preferido es RMG que da origen a ALA, THR, LYS y GLU (pequeño hidrófobo, pequeño hidrófilo, positivo y negativo respectivamente). CYS_7 no se modifica. Se prefiere dejar GLY_5 como está, aunque una proteína homóloga que tiene ALA_5 es tóxica. Las proteínas homólogas que tienen diversos aminoácidos en la posición 9 son tóxicas. Por lo tanto, se usa un codón de variegación NNT que permite $FS^2YCLPARITNVADG$. Se usa NNT en las posiciones 10, 11 y 12 del mismo modo. En la posición 14, después de la cuarta CYS, se permiten ALA, THR, LYS, o GLU (mediante un codón RMG). Esta variegación permite $1{,}053 \cdot 10^7$ secuencias de aminoácidos codificadas por $1{,}68 \cdot 10^7$ secuencias de ADN. Las bibliotecas que tienen $2{,}0 \cdot 10^7$, $3{,}0 \cdot 10^7$ y $5{,}0 \cdot 10^7$ transformantes independientes presentarán, respectivamente, -70%, -83% y -95% de las secuencias permitidas. También son adecuadas otras variegaciones. En referencia a las α-conotoxinas, véase, *inter alia* ALMQ89, CRUZ85, GRAY83, GRAY84 y PARD89.

La micro-proteína parental puede ser una de las proteínas denominadas "Híbrido 1" e "Híbrido 2" por Pease *et al*. (PEAS90); véase Figura 4 de PEAS90. Un conjunto de restos preferidos para modificar cualquier proteína está compuesto por:

Aminoácido Parental	Codón Variegado	Aminoácidos permitidos	Secuencias de AA/secuencias de ADN
A5	RVT	ADGTNS	6/6
P6	VYT	PTALIV	6/6
E7	RRS	$EDNKSRG^2$	7/8
T8	VHG	TPALMVQKE	9/9
A9	VHG	ATPLMVQKE	9/9
A10	RMG	AEKT	4/4
K12	VHG	KQETPALMV	9/9
Q16	NNG	$L^3R^2S.WPQMTKVAEG$	13/15

Esto proporciona $9{,}55 \cdot 10^6$ secuencias de aminoácidos codificadas por $1{,}26 \cdot 10^7$ secuencias de ADN. Una biblioteca que comprende $5{,}0 \cdot 10^7$ transformantes permite la expresión del 98,2% de todas las secuencias posibles. En cada posición se permite el amino parental.

En la posición 5 se proporcionan aminoácidos que son compatibles con un giro. En la posición 6 se permiten ILE y VAL ya que tienen carbonos β ramificados y pueden hacer rígida la cadena. En la posición 7 se permiten ASP, ASN y SER que aparecen a menudo en los extremos amino de las hélices. En las posiciones 8 y 9 se permiten varios aminoácidos que favorecen la formación de la hélice (ALA, LEU, MET, GLN, GLU y LYS) que tienen diferentes cargas e hidrofobicidades ya que son parte de la hélice adecuada. La posición 10 está más allá del borde de la hélice, de modo que se permite un conjunto más pequeño (ALA, THR, LYS y GLU). Este conjunto no solamente incluye 3 aminoácidos que favorecen la formación de la hélice, sino también THR que se tolera bien y que también proporciona hidrofilia positiva, negativa y neutra.

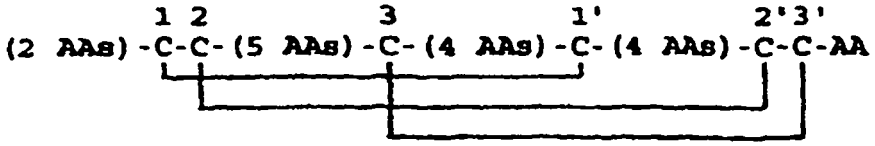
Los grupos laterales de 12 y 16 protegen la misma región que los restos mencionados anteriormente. En estas posiciones, se permite una amplia diversidad de aminoácidos con preferencia por los aminoácidos que favorecen la formación de la hélice.

En lugar de esto, la micro-proteína parental pueden ser un polipéptido compuesto por los restos 9-24 y 31-40 de aprotinina y que posee dos puentes disulfuro (Cys9-Cys22 y Cys14-Cys38). Dicho polipéptido tendría la misma topología de puentes disulfuro que la α-conotoxina y estos dos puentes podrían tener tramos de 12 y 17, respectivamente.

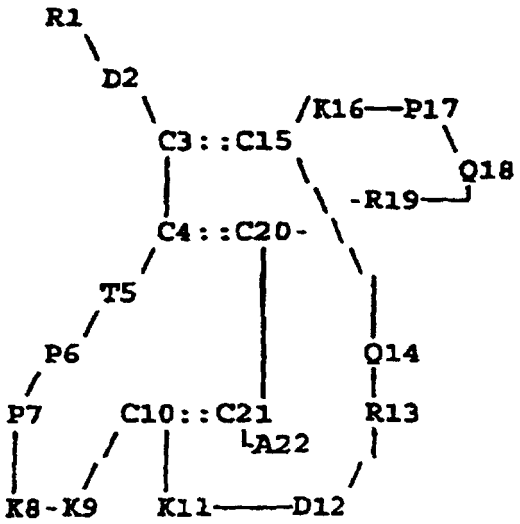
Los restos 23, 24 y 31 se variegan para codificar el conjunto de restos aminoacídicos [G,S,R,D,N,H,P,T,A] de modo que se descubre una secuencia que favorece la formación de un giro de la geometría necesaria. Se usa tripsina o anhidrotripsina como la molécula de afinidad para enriquecer en GP que presenten una micro-proteína que se pliegue en una estructura estable similar a BPTI en la región P1.

Micro-Proteínas Parentales con Tres Puentes Disulfuro

Los caracoles cono (género Conus) producen venenos (conotoxinas) que tienen 10-30 aminoácidos de longitud y son particularmente ricos en puentes disulfuro. Por lo tanto son micro-proteínas arquetípicas. Pueden modelarse nuevas micro-proteínas con tres puentes disulfuro después las μ-conotoxinas (GIIIA, GIIIB, GIIIC) o Ω-(GVIA, GVIB, GVIC, GVIIA, GVIIB, MVIIA, MVIIB, etc.). Las μ-conotoxinas tienen la siguiente estructura conservada:

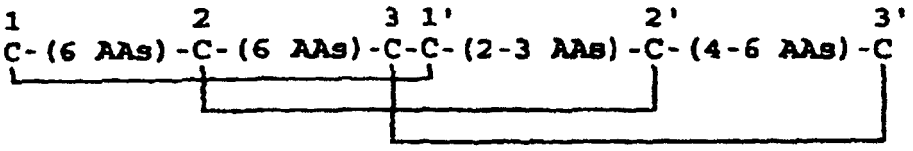


No se ha publicado ninguna estructura 3D de una μ-conotoxina. Hidaka *et al*. (HIDA90) establecieron la conectividad de los puentes disulfuro. El siguiente diagrama representa la geografutoxina I (también conocida como μ-conotoxina GIIIA).



La conexión entre R19 y C20 podría transcurrir por encima o por debajo de la cadena de Q14 a C15. Una forma de variegación preferida es modificar los restos en un bucle. Puesto que el bucle más largo contiene solamente cinco aminoácidos, es apropiado modificar también los restos conectados a las cisteínas que forman el bucle. Por ejemplo, pueden modificarse los restos 5 hasta 9 más 2, 11, 19 y 22. Otra variegación útil sería modificar los restos 11-14 y 16-19 cada uno a través de 8 aminoácidos. En referencia a las μ-conotoxinas, véase BECK89b, BECK89c, CRUZ89 y HIDA90.

Las Ω-conotoxinas pueden representarse de la siguiente manera:



El péptido King Kong tiene la misma disposición de puentes disulfuro que las Ω-conotoxinas, pero una actividad biológica diferente. Woodward *et al*. (WOOD90) presentan las secuencias de tres proteínas homólogas de C. textile. Dentro del dominio de la toxina madura, solamente se conservan las cisteínas. La separación de las cisteínas se conserva de forma exacta, pero ninguna otra posición tiene el mismo aminoácido en las tres secuencias y solamente unas pocas posiciones muestran concordancia de emparejamiento en el plano. De este modo, se concluye que todas las posiciones (excepto las cisteínas) pueden sustituirse libremente con una probabilidad muy alta de que se forme una estructura de puentes disulfuro estables. En referencia a las Ω-conotoxinas véase HILL89 y SUNX87.

Otra micro-proteína que puede usarse como dominio de unión parental es el inhibidor de la tripsina I de *Cucurbita maxima* (CMTI-I); también es adecuado el CMTI-III. Son miembros de la familia de inhibidores de serina proteasa de la calabaza, que también incluyen inhibidores de la calabaza, del calabacín y del pepino (WIEC85). McWherter *et al*.

(MCWH89) describen variantes de secuencia sintética de los inhibidores de la proteasa de la semilla de la calabaza que tienen afinidad por la elastasa y la catepsina G. de los leucocitos humanos. Por supuesto, puede usarse cualquier miembro de esta familia.

CMTI-I es una de las proteínas más pequeñas conocidas, que comprende solamente 29 aminoácidos que se mantienen en conformación fija mediante tres puentes disulfuro. Bode y colegas han estudiado la estructura usando difracción de rayos X (BODE89) y RMN (HOLA89a,b). CMTI-I tiene forma elipsoidal; carece de hélices o de láminas β, pero contiene giros y tramos de polipéptidos cortos de conexión. El emparejamiento de los puentes disulfuro es Cys3-Cys20, Cys10-Cys22 y Cys16-Cys28. En el complejo CMTI-I:tripsina estudiado por Bode *et al*., 13 de los 29 restos inhibidores están en contacto directo con la tripsina; la mayoría de ellos están en el segmento de unión principal Val2 (P4)-Glu9 (P4') que contiene el sitio de unión reactiva Arg5(P1)-Ile6 y está en una conformación que se observa también para otros inhibidores de serina proteinasa.

CMTI-I tiene una K_i para la tripsina de $\sim 1{,}5 \cdot 10^{-12}$ M. McWherter *et al*. sugirieron que la sustitución de "grupos moderadamente hidrófobos que forman una masa" en P1 otorgada especificidad por HEL. Descubrieron que un conjunto de restos más amplio (VAL, ILE, LEU, ALA, PHE, MET y GLY) proporcionaba una unión detectable a HEL. Para la catepsina G, esperaban que se percibieran muy fuertemente los grupos laterales (especialmente aromáticos) en masa. Descubrieron que PRE, LEU, MET y ALA eran funcionales con sus criterios; no ensayaron TRP, TYR, o HIS. (Véase que ALA tiene el segundo grupo lateral *más pequeño* disponible).

Una estrategia de variegación inicial preferida sería modificar algunos o todos los restos ARG_1, VAL_2, PRO_4, ARG_5, ILE_6, LEU_7, MET_8, GLU_9, LYS_{11}, HIS_{25}, GLY_{26}, TYR_{27} y GLY_{29}. Si la diana fuera HNE, por ejemplo, podría sintetizarse ADN que contuviera las siguientes posibilidades:

Parental	Codón vg	Aminoácidos permitidos	Nº de secuencias de AA/Nº de secuencias de ADN
ARG_1	VNT	RSLPHITNVADG	12/12
VAL_2	NWT	VILFYHND	8/8
PRO_4	VYT	PLTIAV	6/6
ARG_5	VNT	RSLPHITNVADG	12/12
ILE6	NNK	los 20	20/31
LEU_7	VWG	LQMKVE	6/6
TYR_{27}	NAS	YHQNKDE.	7/8

Esto permite aproximadamente $5{,}81 \cdot 10^6$ secuencias de aminoácidos codificadas por aproximadamente $1{,}03 \cdot 10^7$ secuencias de ADN. Una biblioteca que comprende $5{,}0 \cdot 10^7$ transformantes independientes proporcionaría ~99% de las secuencias posibles. También podrían usarse otros esquemas de variegación.

Otros inhibidores de esta familia incluyen:

El Inhibidor de la tripsina I de *Citrullus vulgaris* (OTLE87),

El Inhibidor de la tripsina II de *Bryonia dioica* (OTLE87),

El Inhibidor de la tripsina I de *Cucurbita maxima* (en OTLE87),

El Inhibidor de la tripsina III de *Cucurbita maxima* (en OTLE87),

El Inhibidor de la tripsina IV de *Cucurbita maxima* (en OTLE87),

El Inhibidor de la tripsina II de *Cucurbita pepo* (en OTLE87),

El Inhibidor de la tripsina III de *Cucurbita pepo* (en OTLE87),

El Inhibidor de la tripsina IIb de *Cucumis sativus* (en OTLE87),

El Inhibidor de la tripsina IV de *Cucumis sativus* (en OTLE87),

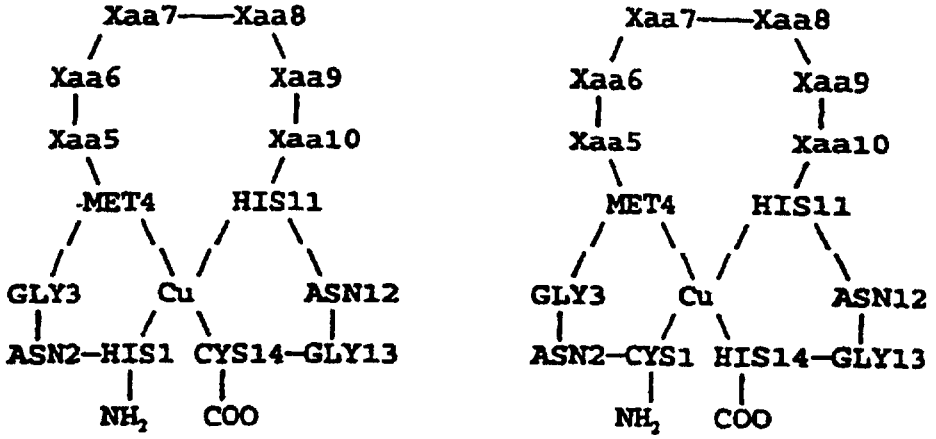
El Inhibidor de la tripsina II de *Ecballium elaterium* (FAVE89), y el inhibidor CM-1 de *Momordica repens* (en OTLE87).

Otras micro-proteínas que pueden usarse como un dominio de unión potencial inicial son las enterotoxinas termo-estables obtenidas de algunas *E. Coli*, *Citrobacter freundii* enterotoxogénicas (GUAR89). Se sabe que *E. coli* secreta estas micro-proteínas y que son extremadamente estables. Los trabajos relacionados con la síntesis, clonación, expresión y propiedades de estas proteínas incluyen BHAT86, SEKI85, SHIM87, TAKA85, TAKE90, THOM85a,b, YOSH85, DALL90, DWAR89, GARI87, GUZM89, GUZM90, HOUG84, KUBO89, KUPE90, OKAM87, OKAM88 y OKAM90.

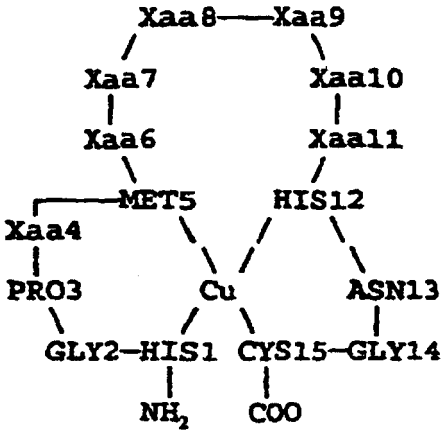
Ejemplo comparativo IV

Una mini-proteína que tiene una reticulación compuesta por Cu(II), una cisteína, dos histidinas y una metionina

Es probable que secuencias tales como HIS-ASN-GLY-MET-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-HIS-ASN-GLY-CYS y CYS-ASN-GLY-MET-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-HIS-ASN-GLY-HIS, se combinen con Cu(II) para formar estructuras como se muestran en el diagrama:



Es probable que otras disposiciones de HIS, MET, HIS y CYS a lo largo de la cadena también formen estructuras similares. Los aminoácidos ASN-GLY en las posiciones 2 y 3 y en las posiciones 12 y 13 proporcionan los aminoácidos que llevan los ligandos de unión a metales con suficiente flexibilidad para que se reúnan y se unan al metal. Pueden usarse otras secuencias de conexión, por ejemplo, GLY-ASN, SER-GLY, GLY-PRO, GLY-PRO-GLY o PRO-GLY-ASN. También es posible modificar uno o más restos en los bucles que unen el primer y el segundo o el tercer y cuarto restos de unión a metales. Por ejemplo,



es probable que forme la estructura representada para una gran diversidad de aminoácidos en Xaa4. Se espera que los grupos laterales de Xaa4 y Xaa6 estén muy cerca entre sí y en la superficie de la mini-proteína.

Los aminoácidos variables se mantienen de modo que tengan una flexibilidad limitada. Esta reticulación tiene algunas diferencias con el enlace disulfuro. La separación entre $C_{\alpha 4}$ y $C_{\alpha 11}$ es mayor que la separación de los C_{α} de una cisteína. Además, se espera que la interacción de los restos 1 a 4 y 11 a 14 con el ión metálico limite la movilidad de los restos 5 a 10 más que un puente disulfuro entre los restos 4 y 11. Un único puente disulfuro ejerce una fuerte limitación de separación sobre los carbonos α de los residuos unidos, pero muy poca limitación direccional sobre, por ejemplo, el vector de N a C en la cadena principal.

Para la secuencia deseada, los grupos laterales de los restos 5 hasta 10 pueden formar interacciones específicas con la diana. Otras cantidades de aminoácidos variables, por ejemplo 4, 5, 7 ó 3, son adecuadas. Pueden usarse tramos mayores cuando la secuencia encerrada contiene segmentos que tienen un alto potencial para formar α-hélices u otra estructura secundaria que limite la libertad conformacional de la cadena principal del polipéptido. Aunque una mini-proteína que tiene cuatro CYS podría formar tres emparejamientos distintos, una mini-proteína que tiene dos HIS, una MET y una CYS puede formar solamente dos complejos distintos con el Cu. Estas dos estructuras se relacionan mediante simetría especular a través del Cu. Como las dos HIS son distinguibles, las estructuras son diferentes.

Cuando dichas mini-proteínas que contienen metal se presentan en un fago filamentoso, las células que producen el fago pueden cultivarse en presencia del ión metálico apropiado, o el fago puede exponerse al ión metálico después de que se separe de las células.

Ejemplo comparativo 5

Una mini-proteína que tiene una reticulación compuesta por Zn(II) y cuatro cisteínas

Se ejemplifica una reticulación similar a la que se muestra en el Ejemplo XV mediante las proteínas de dedo de cinc (GIBS8B, GAUS87, PARR88, FRAN87, CHOW87, HARD90). Una familia de dedos de cinc tiene dos restos CYS y dos restos HIS en posiciones conservadas que se unen al Zn++ (PARR88, FRAN87, CHOW87, EVAN88, BERG88, CHAV88). Gibson *et al.* (GIBS88) revisan varias secuencias que se piensan forman dedos de zinc y proponen un modelo tridimensional para estos compuestos. La mayoría de estas secuencias tienen dos CYS y dos restos HIS en posiciones conservadas, pero algunas tienen tres restos CYS y un HIS. Gauss *et al.* (GAUS87) también presentan una proteína de dedo de cinc que tienen tres restos CYS y un HIS que se unen al cinc. Hard *et al.* (HARD90) presentan la estructura 3D de una proteína que comprende dos dedos de cinc, cada uno de los cuales tiene cuatro restos CYS. Todas estas proteínas de unión al cinc son estables en el entorno intracelular reductor.

Un ejemplo preferido de una mini-proteína reticulada CYS::cinc comprende los restos 440 a 461 de la secuencia que se muestra en la Figura 1 de HARD90. Los restos 444 hasta 456 pueden variegarse. Una de dichas variegaciones es de la siguiente manera:

Parental	Permitido	Nº de AA/Nº de ADN
SER444	SER, ALA	2/2
ASP445	ASP, ASN, GLU, LYS	4/4
GLU446	GLU, LYS, GLN	3/3
ALA447	ALA, THR, GLY, SER	4/4
SER448	SER, ALA	2/2
GLY449	GLY, SER, ASN, ASP	4/4
CYS450	CYS, PHE, ARG, LEU	4/4
HIS451	HIS, GLN, ASN, LYS, ASP, GLU	6/6
TYR452	TYR, PHE, HIS, LEU	4/4
GLY453	GLY, SER, ASN, ASP	4/4
VAL454	VAL, ALA, ASP, GLY, SER, ASN,THR, ILE	8/8
LEU455	LEU, HIS, ASP, VAL	4/4
THR456	THR, ILE, ASN, SER	4/4

Esto conduce a $3{,}77 \cdot 10^7$ secuencias de ADN que codifican la misma cantidad de secuencias de aminoácidos. Una biblioteca que tienen $1{,}0 \cdot 10^8$ transformantes independientes presentará el 93% de las secuencias permitidas; $2{,}0 \cdot 10^8$ transformantes independientes presentarán el 99,5% de las secuencias permitidas.

TABLA 2

Proteínas de la Superficie Externa Preferidas

Paquete Genético	Proteína de la Superficie Externa Preferida	Razones para la Preferencia
M13	proteína de cubierta	a) extremo amino expuesto, (gpVIII) b) procesamiento postraduccional precedible c) numerosas copias en el virión d) datos de la fusión disponibles
	gp III	a) datos de la fusión disponibles b) extremo amino expuesto c) ejemplo de trabajo disponible
PhiX174	Proteína G	a) se sabe que está en el exterior del virión, b) lo suficientemente pequeña para que el gen G-ipbd pueda sustituir al gen H
E. coli	LamB	a) datos de la fusión disponibles, b) no esencial
	OmpC	a) modelo topológico b) no esencial; abundante
	OmpA	a) modelo topológico. b) no esencial; abundante c) homólogos en otros géneros O m p F a) modelo topológico b) no esencial: abundante PhoEa) modelo topológico b) no esencial: abundante c) inducible
B. subtilis	CotC	a) no hay procesamiento de esporas postraduccional, b) secuencia distintiva que provoca que la proteína se sitúe en la cubierta de la espora c) no esencial
	CotD	igual que para CotC

TABLA 10

Abundancia obtenida a partir de diversos codones vg

A. Codón fxS optimizado, limitado por [D] + [E] = [k] + [R]

	T	C	A	G	
1	0,26	0,18	0,26	0,30	f
2	0,22	0,16	0,40	0,22	x
3	0,5	0,0	0,0	0,5	S

Aminoácido	Abundancia		Aminoácido	Abundancia	
A	4,80%		C	2,86%	
D	6,00%		E	6,00%	
F	2,86%		G	6,60%	
H	3,60%		I	2,86%	
K	5,20%		L	6,82%	
M	2,86%		N	5,20%	
P	2,88%		Q	3,60%	
R	6,82%		S	7,02%	mfaa
T	4,16%		V	6,60%	
W	2,86%	lfaa	Y	5,20%	
Detención	5,20%				

[D] + [E] = [K] + [R] = 0,12

proporción = Abun(W)/Abun(S) = 0,4074

j	(1/proporción)j	(proporción)j	sin
1	2,454	0,4074	0,9480
2	6,025	0,1660	0,8987
3	14,788	0,0676	0,8520
4	36,298	0,0275	0,8077
5	89,095	0,0112	0,7657
6	218,7	$4,57 \cdot 10^{-3}$	0,7258
7	536,8	$1,86 \cdot 10^{-3}$	0,6881

TABLA 10 (continuación)

B. Sin limitar, optimizado

	T	C	A	G
1	0,27	0,19	0,27	0,27
2	0,21	0,15	0,43	0,21
3	0,5	0,0	0,0	0,5

Aminoácido	Abundancia	Aminoácido	Abundancia
A	4,05%	C	2,84%
D	5,81%	E	5,81%
F	2,84%	G	5,67%
H	4,08%	I	2,84%
K	5,81%	L	6,83%
M	2,84%	N	5,81%
P	2,85%	Q	4,08%
R	6,83%	S	6,89% mfaa
T	4,05%	V	5,67%
W	2,84% lfaa	Y	5,81%
detención	5,81%		

[D] + [E] = 0,1162 [K] + [R] = 0,1264

proporción = Abun(P)/Abun(S) = 0,41176

i	(1/proporción)j	(proporción)j	sin detención
1	2,4286	0,41176	0,9419
2	5,8981	0,16955	0,8872
3	14,3241	0,06981	0,8356
4	34,7875	0,02875	0,7871
5	84,4849	0,011836	0,74135
6	205,180	0,004874	0,69828
7	498,3	$2,007 \cdot 10^{-3}$	0,6577

TABLA 10 (continuación)

c. NNT optimizado

	T	C	A	G
1	0,2071	0,2929	0,2071	0,2929
2	0,2929	0,2071	0,2929	0,2071
3	1,0	0,0	0,0	0,0

Aminoácido	Abundancia	Aminoácido	Abundancia
A	6,06%	C	4,29% lfaa
D	8,58%	E	ninguno
F	6,06%	G	6,06%
H	8,58%	I	6,06%
K	ninguno	L	8,58%
M	ninguno	N	6,06%
P	6,06%	Q	ninguno
R	6,06%	S	8,58% mfaa
T	4,29% lfaa	V	8,58%
W	ninguno	Y	6,06%
detención	ninguno		

i	(1/proporción)j	(proporción)j	sin detención
1	2,0	0,5	1,0
2	4,0	0,25	1,0
3	8,0	0,125	1,0
4	16,0	0,0625	1,0
5	32,0	0,03125	1,0
6	64,0	0,015625	1,0
7	128,0	0,078125	1,0

TABLA 10 (continuación)

D. NNG optimizado

	T	C	A	G
1	0,23	0,21	0,23	0,33
2	0,215	0,285	0,285	0,215
3	0,0	0,0	0,0	1,0

Aminoácido	Abundancia	Aminoácido	Abundancia
A	9,40%	C	ninguno
D	ninguno	E	9,40%
F	ninguno	G	7,10%
H	ninguno	I	ninguno
K	6,60%	L	9,50% mfaa
M	4,90%	N	ninguno
P	6,00%	Q	6,00%
R	9,50%	S	6,60%
T	6,6 %	V	7,10%
W	4,90% lfaa	Y	ninguno
detención	6,60%		

j	(1/proporción)j	(proporción)j	sin detención
1	1,9388	0,51579	0,934
2	3,7588	0,26604	0,8723
3	7,2876	0,13722	0,8148
4	14,1289	0,07078	0,7610
5	27,3929	$3,65 \cdot 10^{-2}$	0,7108
6	53,109	$1,88 \cdot 10^{-2}$	0,6639
7	102,96	$9,72 \cdot 10^{-3}$	0,6200

TABLA 10 (continuación)

E. NNS sin optimizar (NNK da una distribución idéntica)

	T	C	A	G
1	0,25	0,25	0,25	0,25
2	0,25	0,25	0,25	0,25
3	0,0	0,5	0,0	0,5

Aminoácido	Abundancia	Aminoácido	Abundancia
A	6,25%	C	3,125%
D	3,125%	E	3,125%
F	3,125%	G	6,25%
H	3,125%	I	3,125%
K	3,125%	L	9,375%
M	3,125%	N	3,125%
P	6,25%	Q	3,125%
R	9,375%	S	9,375%
T	6,25%	V	6,25%
W	3,125%	Y	3,125%
detención	3,125%		

i	(1/proporción)j	(proporción)j	sin detención
1	3,0	0,33333	0,96875
2	9,0	0,11111	0,9385
3	27,0	0,03704	0,90915
4	81,0	0,01234567	0,8807
5	243,0	0,0041152	0,8532
6	729,0	$1,37 \cdot 10^{-3}$	0,82655
7	2187,0	$4,57 \cdot 10^{-4}$	0,8007

TABLA 50

IPBD Ejemplares

IPBD	Cantidad de aminoácidos	Estructura	Reticulaciones	Secretada	Organismo fuente	AfM
CMTI-III	26	RMN	3 SS	Si	pepino	Tripsina
ST-I_A	13	RMN	3 SS	Si	E. coli	MAbs & guanilato ciclasa
α-Conotoxinas	13-15	RMN	2 SS	Si	Caracoles Conus	Receptor

μ-Conotoxinas	20-25	RMN	3 SS	Si	Caracoles Conus	Receptor
Ω-Conotoxinas	25-30	RMN	3 SS	Si	Caracoles Conus	Receptor
Péptidos King-Kong	25-30	-	3 SS	Si	Caracoles Conus	Mab
Caribdotoxina (toxina de escorpión)	37	RMN	3 SS 7-28, 3-33 17-35(1: 4, 2:5, 3:6)	Si	Leiurus quinquestriatus hebraeus	Canal K^+ dependiente de Ca^{2+}
Apamina (veneno de abeja)	12	RMN	2 SS (1:3, 2:4)	Si	Abejas	Mab, Receptor (¿)

TABLA 130

Muestreo de una Biblioteca codificada por (NNK)^6

A. Cantidad de hexapéptidos en cada clase

total = 64,000,000 de secuencias sin detención

α puede ser uno de [WMFYCIKDENHQ]
Φ puede ser uno de [PTAVG]
Ω puede ser uno de [SLR]

αααααα	=	2985984.	Φαααααα	=	7464960.
Ωαααααα	=	4478976.	ΦΦααααα	=	7776000.
ΦΩαααααα	=	9331200.	ΩΩαααααα	=	2799360.
ΦΦΦαααα	=	4320000.	ΦΦΩαααα	=	7776000.
ΦΩΩαααα	=	4665600.	ΩΩΩαααα	=	933120.
ΦΦΦΦααα	=	1350000.	ΦΦΦΩααα	=	3240000.
ΦΦΩΩααα	=	2916000.	ΦΩΩΩααα	=	1166400.
ΩΩΩΩααα	=	174960.	ΦΦΦΦΦα	=	225000.
ΦΦΦΦΩα	=	675000.	ΦΦΦΩΩα	=	810000.
ΦΦΩΩΩα	=	486000.	ΦΩΩΩΩα	=	145800.
ΩΩΩΩΩα	=	17496.	ΦΦΦΦΦΦ	=	15625.
ΦΦΦΦΦΩ	=	56250.	ΦΦΦΦΩΩ	=	84375.
ΦΦΦΩΩΩ	=	67500.	ΦΦΩΩΩΩ	=	30375.
ΦΩΩΩΩΩ	=	7290.	ΩΩΩΩΩΩ	=	729.

ΦΦΩΩαα, por ejemplo, representa el conjunto de péptidos que tienen dos aminiácidos de la clase α, dos de Φ, y dos de Ω dispuestos en cualquier orden. Hay, por ejemplo, $729 = 3^6$ secuencias compuestas completamente de S, L y R.

TABLA 130 (continuación)

B. Probabilidad de que cualquier secuencia de ADN sin detención codifique un hexapéptido de la clase indicada

	P	% de la clase
αααααα...	3,364E-03	(1,13E-07)
Φαααααα...	1,682E-02	(2,25E-07)
Ωαααααα...	1,514E-02	(3,38E-07)
ΦΦαααααα...	3,505E-02	(4,51E-07)
ΦΩαααααα...	6,308E-02	(6,76E-07)
ΩΩαααααα...	2,839E-02	(1,01E-06)
ΦΦΦαααα...	3,894E-02	(9,01E-07)
ΦΦΩαααα...	1,051E-01	(1,35E-06)
ΦΩΩαααα...	9,463E-02	(2,03E-06)
ΩΩΩαααα...	2,839E-02	(3,04E-06)
ΦΦΦΦααα...	2,434E-02	(1,80E-06)
ΦΦΦΩααα...	8,762E-02	(2,70E-06)
ΦΦΩΩααα...	1,183E-01	(4,06E-06)
ΦΩΩΩααα...	7,097E-02	(6,08E-06)
ΩΩΩΩααα...	1,597E-02	(9,13E-06)
ΦΦΦΦΦα...	8,113E-03	(3,61E-06)
ΦΦΦΦΩα...	3,651E-02	(5,41E-06)
ΦΦΦΩΩα...	6,571E-02	(8,11E-06)
ΦΦΩΩΩα...	5,914E-02	(1,22E-05)
ΦΩΩΩΩα...	2,661E-02	(1,83E-05)
ΩΩΩΩΩα...	4,790E-03	(2,74E-05)
ΦΦΦΦΦΦ...	1,127E-03	(7,21E-06)
ΦΦΦΦΦΩ...	6,084E-03	(1,08E-05)
ΦΦΦΦΩΩ...	1,369E-02	(1,62E-05)
ΦΦΦΩΩΩ...	1,643E-02	(2,43E-05)
ΦΦΩΩΩΩ...	1,109E-02	(3,65E-05)
ΦΩΩΩΩΩ...	3,992E-03	(5,48E-05)
ΩΩΩΩΩΩ...	5,988E-04	(8,21E-05)

TABLA 130 (continuación)

C. Cantidad de secuencias de aminoácidos diferentes en cada clase esperada para diversos tamaños de biblioteca

Tamaño de biblioteca = 1,0000E+06

total = 9,7446E+05 % muestreado 1,52

Clase	Cantidad	%	Clase	Cantidad	%
αααααα...	3362,6	0,1	Φαααααα...	16803,4	0,2
Ωαααααα...	15114,6	0,3	ΦΦαααααα...	34967,8	0,4
ΦΩαααααα...	62871,1	0,7	ΩΩαααααα...	28244,3	1,0
ΦΦΦαααα...	38765,7	0,9	ΦΦΩαααα...	104432,2	1,3
ΦΩΩαααα...	93672,7	2,0	ΩΩΩαααα...	27960,3	3,0
ΦΦΦΦααα...	24119,9	1,8	ΦΦΦΩααα...	86442,5	2,7
ΦΦΩΩααα...	115915,5	4,0	ΦΩΩΩααα...	68853,5	5,9
ΩΩΩΩααα...	15261,1	8,7	ΦΦΦΦΦα...	7968,1	3,5
ΦΦΦΦΩα...	35537,2	5,3	ΦΦΦΩΩα...	63117,5	7,8
ΦΦΩΩΩα...	55684,4	11,5	ΦΩΩΩΩα...	24325,9	16,7
ΩΩΩΩΩα...	4190,6	24,0	ΦΦΦΦΦΦ...	1087,1	7,0
ΦΦΦΦΩ...	5767,0	10,3	ΦΦΦΦΩΩ...	12637,2	15,0
ΦΦΦΩΩΩ...	14581,7	21,6	ΦΦΩΩΩΩ...	9290,2	30,6
ΦΩΩΩΩΩ...	3073,9	42,2	ΩΩΩΩΩΩ...	408,4	56,0

Tamaño de biblioteca = 3.0000E+06

total = 2,7885E+06 % muestreado 4,36

αααααα...	10076,4	0,3	Φαααααα...	50296,9	0,7
Ωαααααα...	45190,9	1,0	ΦΦαααααα...	104432,2	1,3
ΦΩαααααα...	187345,5	2,0	ΩΩαααααα...	83880,9	3,0
ΦΦΦαααα...	115256,6	2,7	ΦΦΩαααα...	309107,9	4,0
ΦΩΩαααα...	275413,9	5,9	ΩΩΩαααα...	81392,5	8,7
ΦΦΦΦααα...	71074,5	5,3	ΦΦΦΩααα...	252470,2	7,8
ΦΦΩΩααα...	334106,2	11,5	ΦΩΩΩααα...	194606,9	16,7
ΩΩΩΩααα...	41905,9	24,0	ΦΦΦΦΦα...	23067,8	10,3
ΦΦΦΦΩα...	101097,3	15,0	ΦΦΦΩΩα...	174981,0	21,6
ΦΦΩΩΩα...	148643,7	30,6	ΦΩΩΩΩα...	61478,9	42,2
ΩΩΩΩΩα...	9801,0	56,0	ΦΦΦΦΦΦ...	3039,6	19,5
ΦΦΦΦΦΩ...	15587,7	27,7	ΦΦΦΦΩΩ...	32516,8	38,5
ΦΦΦΩΩΩ...	34975,6	51,8	ΦΦΩΩΩΩ...	20215,5	66,6
ΦΩΩΩΩΩ...	5879,9	80,7	ΩΩΩΩΩΩ...	667,0	91,5

TABLA 130 (continuación)

Tamaño de biblioteca = 1,0000E+07

total = 8,1204E+06 % muestreado 12,69

αααααα...	33455,9(1,1)	Φαααααα...	166342,4(2,2)
Ωαααααα...	148871,1(3,3)	ΦΦαααααα...	342685,7(4,4)
ΦΩαααααα...	609987,6(6,5)	ΩΩαααααα...	269958,3(9,6)
ΦΦΦαααα...	372371,8(8,6)	ΦΦΩαααα...	983416,4(12,6)
ΦΩΩαααα...	856471,6(18,4)	ΩΩΩαααα...	244761,5(26,2)
ΦΦΦΦααα...	222702,0(16,5)	ΦΦΦΩααα...	767692,5(23,7)
ΦΦΩΩααα...	972324,6(33,3)	ΦΩΩΩααα...	531651,3(45,6)
ΩΩΩΩααα...	104722,3(59,9)	ΦΦΦΦΦα...	68111,0(30,3)
ΦΦΦΦΩα...	281976,3(41,8)	ΦΦΦΩΩα...	450120,2(55,6)
ΦΦΩΩΩα...	342072,1(70,4)	ΦΩΩΩΩα...	122302,6(83,9)
ΩΩΩΩΩα...	16364,0(93,5)	ΦΦΦΦΦΦ...	8028,0(51,4)
ΦΦΦΦΦΩ...	37179,9(66,1)	ΦΦΦΦΩΩ...	67719,5(80,3)
ΦΦΦΩΩΩ...	61580,0(91,2)	ΦΦΩΩΩΩ...	29586,1(97,4)
ΦΩΩΩΩΩ...	7259,5(99,6)	ΩΩΩΩΩΩ...	728,8(100,0)

Tamaño de biblioteca = 3,0000E+07

total = 1,8633E+07 % muestreado 29,11

αααααα...	99247,4(3,3)	Φαααααα...	487990,0(6,5)
Ωαααααα...	431933,3(9,6)	ΦΦαααααα...	983416,5(12,6)
ΦΩαααααα...	1712943,0(18,4)	ΩΩαααααα...	734284,6(26,2)
ΦΦΦαααα...	1023590,0(23,7)	ΦΦΩαααα...	2592866,0(33,3)
ΦΩΩαααα...	2126605,0(45,6)	ΩΩΩαααα...	558519,0(59,9)
ΦΦΦΦααα...	563952,6(41,8)	ΦΦΦΩααα...	1800481,0(55,6)
ΦΦΩΩααα...	2052433,0(70,4)	ΦΩΩΩααα...	978420,5(83,9)
ΩΩΩΩααα...	163640,3(93,5)	ΦΦΦΦΦα...	148719,7(66,1)
ΦΦΦΦΩα...	541755,7(80,3)	ΦΦΦΩΩα...	738960,1(91,2)
ΦΦΩΩΩα...	473377,0(97,4)	ΦΩΩΩΩα...	145189,7(99,6)
ΩΩΩΩΩα...	17491,3(100,0)	ΦΦΦΦΦΦ...	13829,1(88,5)
ΦΦΦΦΦΩ...	54058,1(96,1)	ΦΦΦΦΩΩ...	83726,0(99,2)
ΦΦΦΩΩΩ...	67454,5(99,9)	ΦΦΩΩΩΩ...	30374,5(100,0)
ΦΩΩΩΩΩ...	7290,0(100,0)	ΩΩΩΩΩΩ...	729,0(100,0)

TABLA 130 (continuación)

Tamaño de biblioteca = 7,6000E+07

total = 3,2125E+07 % muestreado 50,19

αααααα...	245057,8(8,2)	Φαααααα...	1175010,0(15,7)
Ωαααααα...	1014733,0(22,7)	ΦΦαααααα...	2255280,0(29,0)
ΦΩαααααα...	3749112,0(40,2)	ΩΩαααααα...	1504128,0(53,7)
ΦΦΦαααα...	2142478,0(49,6)	ΦΦΩαααα...	4993247,0(64,2)
ΦΩΩαααα...	3666785,0(78,6)	ΩΩΩαααα...	840691,9(90,1)
ΦΦΦΦααα...	1007002,0(74,6)	ΦΦΦΩααα...	2825063,0(87,2)
ΦΦΩΩααα...	2782358,0(95,4)	ΦΩΩΩααα...	1154956,0(99,0)
ΩΩΩΩααα...	174790,0(99,9)	ΦΦΦΦΦα...	210475,6(93,5)
ΦΦΦΦΩα...	663929,3(98,4)	ΦΦΦΩΩα...	808298,6(99,8)
ΦΦΩΩΩα...	485953,2(100,0)	ΦΩΩΩΩα...	145799,9(100,0)
ΩΩΩΩΩα...	17496,0(100,0)	ΦΦΦΦΦ...	15559,9(99,6)
ΦΦΦΦΩ...	56234,9(100,0)	ΦΦΦΩΩ...	84374,6(100,0)
ΦΦΦΩΩΩ...	67500,0(100,0)	ΦΦΩΩΩ...	30375,0(100,0)
ΦΩΩΩΩΩ...	7290,0(100,0)	ΩΩΩΩΩ...	729,0(100,0)

Tamaño de biblioteca = 1,0000E+08

total = 3,6537E+07 % muestreado 57,09

αααααα...	318185,1(10,7)	Φαααααα...	1506161,0(20,2)
Ωαααααα...	1284677,0(28,7)	ΦΦαααααα...	2821285,0(36,3)
ΦΩαααααα...	4585163,0(49,1)	ΩΩαααααα...	1783932,0(63,7)
ΦΦΦαααα...	2566085,0(59,4)	ΦΦΩαααα...	5764391,0(74,1)
ΦΩΩαααα...	4051713,0(86,8)	ΩΩΩαααα...	888584,3(95,2)
ΦΦΦΦααα...	1127473,0(83,5)	ΦΦΦΩααα...	3023170,0(93,3)
ΦΦΩΩααα...	2865517,0(98,3)	ΦΩΩΩααα...	1163743,0(99,8)
ΩΩΩΩααα...	174941,0(100,0)	ΦΦΦΦΦα...	218886,6(97,3)
ΦΦΦΦΩα...	671976,9(99,6)	ΦΦΦΩΩα...	809757,3(100,0)
ΦΦΩΩΩα...	485997,5(100,0)	ΦΩΩΩΩα...	145800,0(100,0)
ΩΩΩΩΩα...	17496,0(100,0)	ΦΦΦΦΦ...	15613,5(99,9)
ΦΦΦΦΦΩ...	56248,9(100,0)	ΦΦΦΩΩ...	84375,0(100,0)
ΦΦΦΩΩΩ...	67500,0(100,0)	ΦΦΩΩΩ...	30375,0(100,0)
ΦΩΩΩΩΩ...	7290,0(100,0)	ΩΩΩΩΩ...	729,0(100,0)

TABLA 130 (continuación)

Tamaño de biblioteca = 3,0000E+08

total = 5,2634E+07 % muestreado 82,24

αααααα...	856451,3(28,7)	Φααααα...	3668130,0(49,1)
Ωααααα...	2854291,0(63,7)	ΦΦααααα...	5764391,0(74,1)
ΦΩααααα...	8103426,0(86,8)	ΩΩααααα...	2665753,0(95,2)
ΦΦΦααα...	4030893,0(93,3)	ΦΦΩααα...	7641378,0(98,3)
ΦΩΩααα...	4654972,0(99,8)	ΩΩΩααα...	933018,6(100,0)
ΦΦΦΦαα...	1343954,0(99,6)	ΦΦΦΩαα...	3239029,0(100,0)
ΦΦΩΩαα...	2915985,0(100,0)	ΦΩΩΩαα...	1166400,0(100,0)
ΩΩΩΩαα...	174960,0(100,0)	ΦΦΦΦΦα...	224995,5(100,0)
ΦΦΦΦΩα...	674999,9(100,0)	ΦΦΦΩΩα...	810000,0(100,0)
ΦΦΩΩΩα...	486000,0(100,0)	ΦΩΩΩΩα...	145800,0(100,0)
ΩΩΩΩΩα...	17496,0(100,0)	ΦΦΦΦΦ...	15625,0(100,0)
ΦΦΦΦΦΩ...	56250,0(100,0)	ΦΦΦΦΩΩ...	84375,0(100,0)
ΦΦΦΩΩΩ...	67500,0(100,0)	ΦΦΩΩΩΩ...	30375,0(100,0)
ΦΩΩΩΩΩ...	7290,0(100,0)	ΩΩΩΩΩΩ...	729,0(100,0)

Tamaño de biblioteca = 1,0000E+09

total = 6,1999E+07 % muestreado 96,87

αααααα...	2018278,0(67,6)	Φααααα...	6680917,0(89,5)
Ωααααα...	4326519,0(96,6)	ΦΦααααα...	7690221,0(98,9)
ΦΩααααα...	9320389,0(99,9)	ΩΩααααα...	2799250,0(100,0)
ΦΦΦααα...	4319475,0(100,0)	ΦΦΩααα...	7775990,0(100,0)
ΦΩΩααα...	4665600,0(100,0)	ΩΩΩααα...	933120,0(100,0)
ΦΦΦΦαα...	1350000,0(100,0)	ΦΦΦΩαα...	3240000,0(100,0)
ΦΦΩΩαα...	2916000,0(100,0)	ΦΩΩΩαα...	1166400,0(100,0)
ΩΩΩΩαα...	174960,0(100,0)	ΦΦΦΦΦα...	225000,0(100,0)
ΦΦΦΦΩα...	675000,0(100,0)	ΦΦΦΩΩα...	810000,0(100,0)
ΦΦΩΩΩα...	486000,0(100,0)	ΦΩΩΩΩα...	145800,0(100,0)
ΩΩΩΩΩα...	17496,0(100,0)	ΦΦΦΦΦ...	15625,0(100,0)
ΦΦΦΦΦΩ...	56250,0(100,0)	ΦΦΦΦΩΩ...	84375,0(100,0)
ΦΦΦΩΩΩ...	67500,0(100,0)	ΦΦΩΩΩΩ...	30375,0(100,0)
ΦΩΩΩΩΩ...	7290,0(100,0)	ΩΩΩΩΩΩ...	729,0(100,0)

Tamaño de biblioteca = 3,0000E+09

total = 6,3890E+07 % muestreado 99,83

αααααα...	2884346,0(96,6)	Φααααα...	7456311,0(99,9)
Ωααααα...	4478800,0(100,0)	ΦΦααααα...	7775990,0(100,0)
ΦΩααααα...	9331200,0(100,0)	ΩΩααααα...	2799360,0(100,0)
ΦΦΦααα...	4320000,0(100,0)	ΦΦΩααα...	7776000,0(100,0)
ΦΩΩααα...	4665600,0(100,0)	ΩΩΩααα...	933120,0(100,0)
ΦΦΦΦαα...	1350000,0(100,0)	ΦΦΦΩαα...	3240000,0(100,0)
ΦΦΩΩαα...	2916000,0(100,0)	ΦΩΩΩαα...	1166400,0(100,0)
ΩΩΩΩαα...	174960,0(100,0)	ΦΦΦΦΦα...	225000,0(100,0)
ΦΦΦΦΩα...	675000,0(100,0)	ΦΦΦΩΩα...	810000,0(100,0)
ΦΦΩΩΩα...	486000,0(100,0)	ΦΩΩΩΩα...	145800,0(100,0)
ΩΩΩΩΩα...	17496,0(100,0)	ΦΦΦΦΦ...	15625,0(100,0)
ΦΦΦΦΦΩ...	56250,0(100,0)	ΦΦΦΦΩΩ...	84375,0(100,0)
ΦΦΦΩΩΩ...	67500,0(100,0)	ΦΦΩΩΩΩ...	30375,0(100,0)
ΦΩΩΩΩΩ...	7290,0(100,0)	ΩΩΩΩΩΩ...	729,0(100,0)

TABLA 130 (continuación)

D. Fórmulas para las cantidades tabuladas

Ltmn es la cantidad de transformantes independientes.
31**6 es 31 a la sexta potencia; 6*3 significa 6 veces 3.
A = Ltmn/(31**6)

```
..  puede ser uno de  -.
α   puede ser uno de  [WMFYCIKDENHQ.]
Φ   puede ser uno de  [PTAVG]
Ω   puede ser uno de  [SLR]
F0 = (12)**6      F1 = (12)**5      F2 = (12)**4
F3 = (12)**3      F4 = (12)**2      F5 = (12)
F6 = 1

ααααααα = F0 * (1-exp(-A))
Φαααααα = 6 * 5 * F1 * (1-exp(-2*A))
Ωαααααα = 6 * 3 * F1 * (1-exp(-3*A))
ΦΦααααα = (15) * 5**2 * F2 * (1-exp(-4*A))
ΦΩααααα = (6*5)*5*3 *F2  * (1-exp(-6*A))
ΩΩααααα = (15) * 3**2 * F2 * (1-exp(-9*A))
ΦΦΦαααα = (20)*(5**3) * F3 * (1-exp(-8*A))
ΦΦΩαααα = (60)*(5*5*3)*F3* (1-exp(-12*A))
ΦΩΩαααα = (60)*(5*3*3)*F3*(1-exp(-18*A))
ΩΩΩαααα = (20)*(3)**3*F3*(1-exp(-27*A))
ΦΦΦΦααα = (15)*(5)**4*F4*(1-exp(-16*A))
ΦΦΦΩααα = (60)*(5)**3*3*F4*(1-exp(-24*A))
ΦΦΩΩααα = (90)*(5*5*3*3)*F4*(1-exp(-36*A))
ΦΩΩΩααα = (60)*(5*3*3*3)*F4*(1-exp(-54*A))
ΩΩΩΩααα = (15)*(3)**4 * F4 *(1-exp(-81*A))
ΦΦΦΦΦαα = (6)*(5)**5 * F5 * (1-exp(-32*A))
ΦΦΦΦΩαα = 30*5*5*5*5*3*F5*(1-exp(-48*A))
ΦΦΦΩΩαα = 60*5*5*5*3*3*F5*(1-exp(-72*A))
ΦΦΩΩΩαα = 60*5*5*3*3*3*F5*(1-exp(-108*A))
ΦΩΩΩΩαα = 30*5*3*3*3*3*F5*(1-exp(-162*A))
ΩΩΩΩΩαα = 6*3*3*3*3*3*F5*(1-exp(-243*A))
ΦΦΦΦΦΦ = 5**6 * (1-exp(-64*A))
ΦΦΦΦΦΩ = 6*3*5**5*(1-exp(-96*A))
ΦΦΦΦΩΩ = 15*3*3*5**4*(1-exp(-144*A))
ΦΦΦΩΩΩ = 20*3**3*5**3*(1-exp(-216*A))
ΦΦΩΩΩΩ = 15*3**4*5**2*(1-exp(-324*A))
ΦΩΩΩΩΩ = 6*3**5*5*(1-exp(-486*A))
ΩΩΩΩΩΩ = 3**6*(1-exp(-729*A))
total  = ααααααα + Φαααααα + Ωαααααα + ΦΦααααα + ΦΩααααα +
         ΩΩααααα + ΦΦΦαααα + ΦΦΩαααα + ΦΩΩαααα + ΩΩΩαααα +
         ΦΦΦΦααα + ΦΦΦΩααα + ΦΦΩΩααα + ΦΩΩΩααα + ΩΩΩΩααα +
         ΦΦΦΦΦαα + ΦΦΦΦΩαα + ΦΦΦΩΩαα + ΦΦΩΩΩαα + ΦΩΩΩΩαα +
         ΩΩΩΩΩαα + ΦΦΦΦΦΦ + ΦΦΦΦΦΩ + ΦΦΦΦΩΩ + ΦΦΦΩΩΩ +
         ΦΦΩΩΩΩ + ΦΩΩΩΩΩ + ΩΩΩΩΩΩ
```

TABLA 131

Muestreo de un Biblioteca Codificada por (NNT)^4(NNG)^2

X puede ser F,S,Y,C,L,P,H,R,I,T,N,V,A,D,G

Γ puede ser L², R², S,W,P,Q,M,T,K,V,A,E,G

La biblioteca comprende 8,55 · 10E6 secuencias de aminoácidos; 1,47 · 10E7 secuencias de ADN

Cantidad total de secuencias de aa posibles = 8.555.625

x	LVPTARGFYCHIND
S	S
θ	VPTAGWQMKES
Ω	LR

Las posiciones primera, segunda, quinta y sexta pueden alojar x o S; la posición tercera y cuarta puede alojar θ o Ω. Las secuencias se agrupan por la cantidad de x, S,θ, y Ω.

Por ejemplo xxθΩSS representa:
[xxθΩSS, xSθΩxS, xSθΩSx, SSθΩxx, SxθΩxS, SxθΩSx,
xxΩθSS, xSΩθxS, xSΩθSx, SSΩθxx, SxΩθxS, SxΩθSx]

La siguiente tabla muestra la posibilidad de que cualquier secuencia de ADN particular esté dentro de una de las clases definidas.

Tamaño de biblioteca 1,0 Muestreo = 0,0001%

total.....	1,0000E+00	%muestreado	1,1688E-07
xxθθxx.....	3,1524E-01	xxθΩxx.....	2,2926E-01
xxΩΩxx.....	4,1684E-02	xxθθxS.....	1,8013E-01
xxθΩxS.....	1,3101E-01	xxΩΩxS.....	2,3819E-02
xxθθSS.....	3,8600E-02	xxθΩSS.....	2,8073E-02
xxΩΩSS.....	5,1042E-03	xSθθSS.....	3,6762E-03
xSθΩSS.....	2,6736E-03	xSΩΩSS.....	4,8611E-04
SSθθSS.....	1,3129E-04	SSθΩSS.....	9,5486E-05
SSΩΩSS.....	1,7361E-05		

TABLA 131 (continuación)

La siguiente sección muestra cuantas secuencias de cada clase se experan para bibliotecas de diferentes tamaños

Tamaño de biblioteca = 1,0000E+05

total..... 9.9137E+04 fracción muestreada = 1.1587E-02

Tipo	Cantidad	%	Tipo	Cantidad	%
xx0Θxx.....	31416,9(.7)		xx0Θxx.....	22771,4(1,3)	
xxΩΩxx.....	4112,4(2,7)		xx0ΘxS.....	17891,8(1,3)	
xx0ΩxS.....	12924,6(2,7)		xxΩΩxS.....	2318,5(5,3)	
xx0ΘSS.....	3808,1(2,7)		xx0ΩSS.....	2732,5(5,3)	
xxΩΩSS.....	483,7(10,3)		xS0ΘSS.....	357,8(5,3)	
xSΘΩSS.....	253,4(10,3)		xSΩΩSS.....	43,7(19,5)	
SSΘΘSS.....	12,4(10,3)		SSΘΩSS.....	8,6(19,5)	
SSΩΩSS.....	1,4(35,2)				

Tamaño de biblioteca = 1,0000E+06

total..... 9,2064E+05 fracción muestreada = 1,0761E-01

xx0Θxx.....	304783,9(.6,6)		xx0Θxx.....	214394,0(12,7)	
xxΩΩxx.....	36508,6(23,8)		xx0ΘxS.....	168452,5(12,7)	
xx0ΩxS.....	114741,4(23,8)		xxΩΩxS.....	18383,8(41,9)	
xx0ΘSS.....	33807,7(23,8)		xx0ΩSS.....	21666,6(41,9)	
xxΩΩSS.....	3114,6(66,2)		xS0ΘSS.....	2837,3(41,9)	
xSΘΩSS.....	1631,5(66,2)		xSΩΩSS.....	198,4(88,6)	
SSΘΘSS.....	80,1(66,2)		SSΘΩSS.....	39,0(88,6)	
SSΩΩSS.....	3,9(98,7)				

Tamaño de biblioteca = 3,0000E+06

total..... 2,3880E+06 fracción muestreada = 2,7912E-01

xx0Θxx.....	855709,5(18,4)		xx0Θxx.....	565051,6(33,4)	
xxΩΩxx.....	85564,7(55,7)		xx0ΘxS.....	443969,1(33,4)	
xx0ΩxS.....	268917,8(55,7)		xxΩΩxS.....	35281,3(80,4)	
xx0ΘSS.....	79234,7(55,7)		xx0ΩSS.....	41581,5(80,4)	
xxΩΩSS.....	4522,6(96,1)		xS0ΘSS.....	5445,2(80,4)	
xSΘΩSS.....	2369,0(96,1)		xSΩΩSS.....	223,7(99,9)	
SSΘΘSS.....	116,3(96,1)		SSΘΩSS.....	43,9(99,9)	
SSΩΩSS.....	4,0(100,0)				

TABLA 131 (continuación)

Tamaño de biblioteca = 8,5556E+06

total.....	4,9303E+06	fracción muestreada	= 5,7626E-01
xxθθxx.....	2046301,0(44,0)	xxθΩxx.....	1160645,0(68,7)
xxΩΩxx.....	138575,9(90,2)	xxθθxS.....	911935,6(68,7)
xxθΩxS.....	435524,3(90,2)	xxΩΩxS.....	43480,7(99,0)
xxθθSS.....	128324,1(90,2)	xxθΩSS.....	51245,1(99,0)
xxΩΩSS.....	4703,6(100,0)	xSθθSS.....	6710,7(99,0)
xSθΩSS.....	2463,8(100,0)	xSΩΩSS.....	224,0(100,0)
SSθθSS.....	121,0(100,0)	SSθΩSS.....	44,0(100,0)
SSΩΩSS.....	4,0(100,0)		

Tamaño de biblioteca = 1,0000E+07

total.....	5,3667E+06	fracción muestreada	= 6,2727E-01
xxθθxx.....	2289093,0(49.2)	xxθΩxx.....	1254877,0(74,2)
xxΩΩxx.....	143467,0(93.4)	xxθθxS.....	985974,9(74,2)
xxθΩxS.....	450896,3(93.4)	xxΩΩxS.....	43710,7(99,6)
xxθθSS.....	132853,4(93.4)	xxθΩSS.....	5151̀6,1(99,6)
xxΩΩSS.....	4703,9(100.0)	xSθθSS.....	6746,2(99,6)
xSθΩSS.....	2464,0(100.0)	xSΩΩSS.....	224,0(100,0)
SSθθSS.....	121,0(100.0)	SSθΩSS.....	44,0(100,0)
SSΩΩSS.....	4,0(100.0)		

Tamaño de biblioteca = 3,0000E+07

total.....	7,8961E+06	fracción muestreada	= 9,2291E-01
xxθθxx.....	4040589,0(86,9)	xxθΩxx.....	1661409,0(98,3)
xxΩΩxx.....	153619,1(100,0)	xxθθxS.....	1305393,0(98,3)
xxθΩxS.....	482802,9(100,0)	xxΩΩxS.....	43904,0(100,0)
xxθθSS.....	142254,4(100,0)	xxθΩSS.....	51744,0(100,0)
xxΩΩSS.....	4704,0(100,0)	xSθθSS.....	6776,0(100,0)
xSθΩSS.....	2464,0(100,0)	xSΩΩSS.....	224,0(100,0)
SSθθSS.....	121,0(100,0)	SSθΩSS.....	44,0(100,0)
SSΩΩSS.....	4,0(100,0)		

TABLA 131 (continuación)

Tamaño de biblioteca = 5,0000E+07

total.....	8,3956E+06	fracción muestreada =	9,8130E-01
xx00xx.....	4491779,0(96,6)	xx0Oxx.....	1688387,0(99,9)
xxOOxx.....	153663,8(100,0)	xx00xS.....	1326590,0(99,9)
xx0OxS.....	482943,4(100,0)	xxOOxS.....	43904,0(100,0)
xx00SS.....	142295,8(100,0)	xx0OSS.....	51744,0(100,0)
xxOOSS.....	4704,0(100,0)	xS00SS.....	6776,0(100,0)
xS0OSS.....	2464,0(100,0)	xSOOSS.....	224,0(100,0)
SS00SS.....	121,0(100,0)	SS0OSS.....	44,0(100,0)
SSOOSS.....	4,0(100,0)		

Tamaño de biblioteca = 1,0000E+08

total.....	8,5503E+06	fracción muestreada =	9,9938E-01
xx00xx.....	4643063,0(99,9)	xx0Oxx.....	1690302,0(100,0)
xxOOxx.....	153664,0(100,0)	xx00xS.....	1328094,0(100,0)
xx0OxS.....	482944,0(100,0)	xxOOxS.....	43904,0(100,0)
xx00SS.....	142296,0(100,0)	xx0OSS.....	51744,0(100,0)
xxOOSS.....	4704,0(100,0)	xS00SS.....	6776,0(100,0)
xS0OSS.....	2464,0(100,0)	xSOOSS.....	224,0(100,0)
SS00SS.....	121,0(100,0)	SS0OSS.....	44,0(100,0)
SSOOSS.....	4,0(100,0)		

TABLA 132

Eficacias Relativas de diversos codones de variegación sencillos

	Cantidad de codones		
	5	6	7
Codón vg	Nº de ADN/Nº de AA [Nº de ADN] (Nº de AA)	Nº de ADN/Nº de AA [Nº de ADN] (Nº de AA)	Nº de ADN/Nº de AA [Nº de ADN] (Nº de AA)
NNK suponiendo que la interrupción desaparece	8,95 [2,86·10^7] (3,2·10^6)	13,86 [8,8·7,10^8] (6,4·10^7)	21,49 [2,75·10^{10}] (1,28·10^9)
NNT	1,38 [1,05·106] (7,59·10^5)	1,47 [1,68·10^7] (1,14·10^7)	1,57 [2,68·10^8] (1,71·10^8)
NNG suponiendo que la interrupción desaparece	2,04 [7,59·10^5] (3,7·10^5)	2,36 [1,14·10^6] (4,83·10^6)	2,72 [1,71·10^8] (6,27·10^7)

TABLA 155

Distancia en Å entre carbonos alfa en octapéptidos

Cadena extendida: ángulo de C_α 1-C_α 2-C_α 3 = 138º

	1	2	3	4	5	6	7	8
1	-							
2		3,8	-					
3		7,1	3,8	-				
4	10,7	7,1	3,8	-				
5	14,2	10,7	7,1	3,8	-			
6	17,7	14,1	10,7	7,1	3,8	-		
7	21,2	17,7	14,1	10,6	7,0	3,8	-	
8	24,6	20,9	17,5	13,9	10,6	7,0	3,8	-

Giro inverso entre los restos 4 y 5.

	1	2	3	4	5	6	7	8
1	-							
2		3,8	-					
3		7,1	3,8	-				
4	10,6	7,0	3,8	-				
5	11,6	8,0	6,1	3,8	-			
6	9,0	5,8	5,5	5,6	3,8	-		
7	6,2	4,1	6,3	8,0	7,0	3,8	-	
8	5,8	6,0	9,1	11,6	10,7	7,2	3,8	-

Alfa hélice: ángulo de C_α 1-C_α 2-C_α 3 = 93º

	1	2	3	4	5	6	7	8
1	-							
2		3,8	-					
3		5,5	3,8	-				
4	5,1	5,4	3,8	-				
5	6,6	5,3	5,5	3,8	-			
6	9,3	7,0	5,6	5,5	3,8	-		
7	10,4	9,3	6,9	5,4	5,5	3,8	-	
8	11,3	10,7	9,5	6,8	5,6	5,6	3,8	-

TABLA 156

Distancias entre carbonos alfa en mini-proteínas cerradas de la forma del ciclo disulfuro (CXXXXC)

Distancia mínima

	1	2	3	4	5	6
1	-					
2	3,8	-				
3	5,9	3,8	-			
4	5,6	6,0	3,8	-		
5	4,7	5,9	6,0	3,8	-	
6	4,8	5,3	5,1	5,2	3,8	-

Distancia media

	1	2	3	4	5	6
1	-					
2	3,8	-				
3	6,3	3,8	-			
4	7,5	6,4	3,8	-		
5	7,1	7,5	6,3	3,8	-	
6	5,6	7,5	7,7	6,4	3,8	-

Distancia máxima

	1	2	3	4	5	6
1	-					
2	3,8	-				
3	6,7	3,8	-			
4	9,0	6,9	3,8	-		
5	8,7	8,8	6,8	3,8	-	
6	6,6	9,2	9,1	6,8	3,8	-

TABLA 820

Fago Peptídico

Nombre	Supuesta secuencia peptídica de unión a estreptavidina	Marcador de resistencia a antibióticos
HPQ	A E G P C H P Q F - - C Q S Y I E G R I V - - - - E...	
DEV(F)	A E - P C H P Q Y R L C Q R P L K Q P P P P P P A E...	
Dev(E)	A E - L C H P Q F P R C N L F R K V P P P P P P A E...	
HPQ6	A E G P C H P Q F P R C Y I E G R I V - - - - - - E...	

										1	1	1	1	1	1	1	1	2	2	2	2	2	2
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
-	-	-	-	C	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	E

TABLA 383

Péptidos de unión a estreptavidina limitados por puentes disulfuro

Clon	glu	gly	V	cys	V	V	V	V	cys	V	por	Frecuencia
Nº 2	glu	gly	tyr	cys	his	pro	gln	phe	cys	pro	ser	4
Nº 4	glu	gly	his	cys	his	pro	gln	phe	cys	ser	ser	3
Nº 5	glu	gly	leu	cys	his	pro	gln	phe	cys	gly	ser	2
Nº 8	glu	gly	asp	cys	his	pro	gln	phe	cys	ser	ser	2
Nº 1	glu	gly	asn	cys	his	pro	gln	phe	cys	pro	ser	1
Nº 3	glu	gly	asp	cys	his	pro	gln	phe	cys	arg	ser	1
Nº 13	glu	gly	asp	cys	his	pro	gln	phe	cys	val	ser	1
				cys	his	pro	gln	phe	cys			consenso

TABLA 839

Secuencias Obtenidas mediante Enriquecimiento de BSA

Clon	glu	gly	V	cys	V	V	V	V	cys	V	por	Frecuencia
Nº 21	glu	gly	gly	cys	phe	lys	arg	asn	cys	tyr	ser	1
Nº 22	glu	gly	his	cys	asp	lys	lys	ile	cys	leu	ser	1
Nº 23	glu	gly	phe	cys	his	thr	ala	ala	cys	phe	ser	1
Nº 24	glu	gly	his	cys	tyr	lys	gly	val	cys	ser	ser	1
Nº 25	glu	gly	his	cys	asp	lys	trp	arg	cys	pro	ser	1
Nº 26	glu	gly	ile	cys	tyr	arg	leu	asp	cys	ile	ser	1
Nº 27	glu	gly	gly	cys	phe	pro	trp	his	cys	phe	ser	1
Nº 28	glu	gly	ser	cys	asp	ser	leu	arg	cys	asp	ser	1

No se observó consenso

Citas

ALMA83a:

Almassy, RC, JC **Fontecilla-Camps**, FL **Suddath** y CE **Bugg**, Entrada 1SN3 en Brookhaven Protein Data Bank, (1983).

ALMA83b:

Almassy, RC, JC **Fontecilla-Camps**, FL **Suddath** y CE **Bugg**, *J Mol Biol* (1983), 170:497ff.

ALMQ89:

Almquist, RG, SR **Kadambi**, DM **Yasuda**, FL **Weitl**, WE **Polgar** y LR **Toll**, *Int J Pept Protein Res*, (diciembre de 1989), 34(6)455-62.

BAKE87:

Baker, K, N **Mackman** e IB **Holland**, *Prog Biophys molec Biol* (1987), 49:89-115.

BASS90:

Bass, **Greene** y **Wells**, *Proteins: Structure, Function and Genetics* (1990) 8:309-14.

BECK80:

Beck, E, *Nucl Acid Res* (1980), 8(13)3011-3024.

BECK83:

Beckwith, J y TJ **Silhavy**, *Methods in Enzymology* (1983), 97:3-11.

BECK89b:

Becker, S, E **Atherton**, H **Michel** y RD **Gordon**, *J Protein Chem,* (junio de 1989), 8(3)393-4.

BECK89c:

Becker, S, E **Atherton** y RD **Gordon**, *Fur J Biochem*, (20 de octubre de 1989), 185(1)79-84.

BENS84:

Benson, SA, E **Bremer** y TJ **Silhavy**, *Proc Natl Acad Sci USA* (1984), 81:3830-34.

BENS88:

Benson, SA, JL **Occi**, BA **Sampson**, *J Mol Biol* (1988) 203(4)961-70.

BENZ88a:

Benz, R y K **Bauer**, *Eur J Biochem* (1988¿), 176:1-19.

BENZ88b:

Benz, R, *Ann Rev Microbiol* (1988), 42:359-93.

BERG88:

Berg, JM, *Proc Natl Acad Sci USA* (1988), 85:99-102.

BETT88:

Better, M, CP **Chang**, RR **Robinson** y AH **Horwitz**, *Science* (1988), 240:1041-1043.

BHAT86:

Bhatnagar, PK y JC **Frantz**, *Develop biol Standard* (1986), 63:79-87.

BODE89:

Bode, W, HJ **Greyling**, R **Huber**, J **Otlewski** y T **Wilusz**, *FEBS Lett* (2 de enero de 1989), 242(2)285-92.

BOEK80: {cita de restauración; effects of mutations near SP-1 cleavage site.}

Boeke, JD, M **Russel** y P **Model**, *J Mol Biol* (1980), 144:103-116.

BOEK82: {cita de restauración; effect of glllp on membrane}

Boeke, JD, P **Model** y ND **Zinder**, *Molec y Gen Genet*, (1982), 186:185-192.

CARM90:

Carmel, G, D **Hellstern**, D **Henning** y JW **Coulton**, *J Bacteriol*, (abril de 1990), 172(4)1861-9.

CARU85:

Caruthers, MH, *Science* (1985), 230:281-285.

CARU87:

Caruthers, MH, P **Gottlieb**, LP **Bracco** y L **Cummings**, en *Protein Structure, Folding, and Design 2*, 1987, Ed. D Oxender (New York, AR Liss Inc) p.9ff.

CATR87:

Catron, KM y CA **Schnaitman**, *J Bacteriol* (1987), 169:4327-34.

CHAR86a:

Charbit, A, JC **Boulain**, A **Ryter** y M **Hofnung**, *EMBO J*, (1986), 5(11)3029-37.

CHAR86b:

Charbit, A, J-C **Boulain** y M **Hofnung**, *Comptes Rendu Acad Sci*, Paris, (1986), 302:617-24.

CHAR87:

Charbit, A, E **Sobczak**, ML **Michel**, A Molla, P **Tiollais** y M **Hofnung**, *J Immunol* (1987), 139:1658-64.

CHAR88b:

Charbit, A, A **Molla**, W **Saurin** y M **Hofnung**, *Gene* (1988), 70(1)181-9.

CHAR88c:

Charbit, A, S Van der **Werf**, V **Mimic**, JC **Boulain**, M **Girard** y M **Hofnung**, *Ann Inst Pasteur Microbiol* (1988), 139(1)45-58.

CHAV88:

Chavrier, P, P **Lemaire**, O **Revelant**, R **Bravo** y P **Charnay**, *Molec Cell Biol* (1988), 8(3)1319-26.

CHOU74:

Chou, PY y GD **Fasman**, *Biochemistry* (1974), 13.(2)222-45.

CHOW87:

Chowdhuury, K, U **Deutsch** y P **Gruss**, *Cell* (1987), 48:771-778.

CLEM81:

Clement, JM y M **Hofnung**, *Cell* (1981), 27:507-514.

CLIC88:

Click, EM, GA McDonald y CA Schnaitman; *J Bacteriol* (1988), 170:2005-2011.

CLUN84;

Clune, A, K-S **Lee** y T **Ferenci**, *Biochem and Biophys Res Comm* (1984), 121:34-40.

CRE184:

Creighton, TE, *Proteins: Structures and Molecular Principles*, W H Freeman & Co, New York, 1984.

CREI88:

Creighton, TE, *BioEssays* (1988), 8(2)57-63.

CRUZ85:

Cruz, LJ, WR **Gray**, BM **Olivera**, RD **Zeikus**, L **Kerr**, D **Yoshikami** y E **Moczydlowski**, *J Biol Chem*, (1985), 260(16)9280-8.

CRUZ89:

Cruz, LJ, G **Kupryszewski**, GW **LeCheminant**, WR **Grey**, BM **Oliveria** y J **Rivier**, *Biochem* (1989), 28:3437-3442.

CWIR90:

Cwirla, SE, EA **Peters**, RW **Barrett** y WJ **Dower**, *Proc Natl Acad Sci USA*, (agosto de 1990), 87:6378-6382.

DALL90:

Dallas, WS, *J Bacteriol* (1990), 172(9)5490-93.

DEGE84:

de **Geus**, P, HM **Verheij**, NH **Reigman**, WPM **Hoekstra** y GH de **Haas**, *EMBO J* (1984), 3(8)1799-1802.

DELA88:

de la **Cruz**, VF, AA **Lal** y TF **McCutchan**, *J Biol Chem*, (1988), 263(9)4318-22.

DEVL90:

Devlin, JJ, LC **Panganiban** y PE **Devlin**, *Science*, (27 de julio de 1990), 249:404-406.

DILL87:

Dill, KA, *Protein Engineering* (1987), 1:369-371.

DUCH88:

Duchene, M, A **Schweized**, F **Lottspeich**, G **Krauss**, M **Marget**, K **Vogel**, B-U von **Specht** y H **Domdey**, *J Bacteriol* (1987), 170:155-162.

DULB86:

Dulbecco, R, Patente de Estados Unidos Nº 4.593.002, 3 de junio de 1986.

DWAR89:

Dwarakanath, P, SS **Viswiswariah**, YVBK **Subrahmanyam**, G **Shanthi**, HM **Jagannatha** y TS **Balganesh**, *Gene* (1989), 81:219-226.

EISE85:

Eisenbeis, SJ, MS. **Nasoff**, SA **Noble**, LP **Bracco**, DR **Dodds**, MH **Caruthers**, *Proc Natl Acad Sci USA* (1985), 82:1084-1088.

ELLE88:

Elleman, TC, *Microbiol Rev* (1988), 52(2)233-47.

EVAN88:

Evans, RM y SM **Hollenberg**, *Cell* (1988), 52:1-3.

FAVE89;

Favel, A, D **Le-Nguyen**, MA **Coletti-Previero** y C **Castro**, *Biochem Biophys Res Comm* (1989), 162:79-82.

FERE82a:

Ferenci. T, *Ann Microbiol* (Inst Pasteur) (1982), 133A: 167-169.

FERE82b:

Ferenci. T y K-S **Lee**, *J Mol Biol* (1982), 160:431-444.

FERE83:

Ferenci, T y KS **Lee**, *J Bacteriol* (1983), 154:984-987.

FERE84:

Ferenci. T, *Trends in Biological Science* (1984) Vol. ?:44-48.

FRAN87:

Frankel, AD, JM **Berg** y CO **Pabo**, *Proc Natl Acad Sci USA* (1987), 84:4841-45.

FREU89:

Freudl, R, H **Schwarz**, M **Degen** y U **Henning**, *J Mol Biol* (1989), 205(4)771-5.

GARI87:

Gariepy, J, AK **Judd** y GK **Schoolnik**, *Proc Natl Acad Sci USA* (1987), 84:8907-11.

GAUS87:

Gauss, P, KB **Krassa**, DS **McPheeters**, MA **Nelson** y L **Gold**, *Proc Natl Acad Sci USA* (1987), 84:8515-19.

GETZ88:

Getzoff, ED, HE **Parge**, DE **McRee** y JA **Tainer**, *Rev Infect Dis* (1988), 10(Supl. 2)S296-299.

GIBS88:

Gibson, TJ, JPM **Postma**, RS **Brown** y P **Argos**, *Protein Engineering* (1988), 2(3)209-218.

GRAY83:

Gray, WR, JE **Rivier**, R **Galyean**, LJ **Cruz** y BM **Olivera**, *J Biol Chem*, (1983), 258(20)12247-51.

GRAY84:

Gray, WR, FA **Luque**, R **Galyean**, E **Atherton** y RC **Sheppard**, BL **Stone**, A **Reyes**, J **Alford**, M **Mcintosh**, BM **Olivera** *et al*. *Biochemistry*, (1984), 23(12)2796-802.

GUAR89:

Cuarino, A, R **Giannella** y MR **Thompson**, *Infection and Immunity* (1989), 57(2)649-52.

GUDM89:

Gudmundsdottir, A, PE **Bell**, MD **Lundrigan** y C **Bradbeer** y RJ **Kadner**, *J Bacteriol*, (diciembre de 1989), 171(12)6526-33.

GUSS88:

Guss, JM, EA **Merritt**, RP **Phizackerley**, R **Hedman**, M **Murata**, KO **Hodgson**, HC **Freeman**, *Science* (1988), 241:806-11.

GUZM89:

Guzman-Verduzco, L-M y YM **Kupersztoch**, *Infection and Immunity* (1989), 57(2)645-48.

GUZM90:

Guzman-Verduzco, L-M y YM **Kupersztoch**, *Molec Microbiol* (1990), 4:253-64.

HARD90:

Hard, T, E **Kellenbach**, R **Boelens**, BA **Maler**, K **Dahlman**, LP **Freedman**, J **Carlstedt-Duke**, KR **Yamamoto**, J-A **Gustafsson** y R **Kaptein**, *Science* (13 de julio de 1990), 249:157-60.

HASH85:

Hashimoto, K, S **Uchida**, H **Yoshida**, Y **Nishiuchi**, S **Sakakibara** y K **Yukari**, *Eur J Pharmacol* (1985), 118(3)351-4.

HATA90:

Hatanaka, Y, E **Yoshida**, H **Nakayama** y Y **Kanaoka**, *Chem Pharm Bull* (Tokyo), (enero de 1990), 38:236-8.

HEDE89:

Hedegaard, L y P **Klemm**, *Gene*, (21 de diciembre de 1989), 85(1)115-24.

HEIN87:

Heine, HG, J **Kyngdon** y T **Ferenci**, *Gene* (1987), 53:287-92.

HEIN88:

Heine, HG, G **Francis**, KS **Lee** y T **Ferenci**, *J Bacteriol* (abril de 1988), 170:1730-8.

HIDA90:

Hidaka, Y, K **Sato**, H **Nakamura**, J **Kobayashi**, Y **Ohizumi** y Y **Shimonishi**, *FEBS Lett* (1990), 264(1) 29-32.

HOCJ85:

Ho, C, M **Jasin** y P **Schimmel**, *Science* (1985), 229:389-93.

HOLA89a:

Holak, TA, D **Gondol**, J **Otlewski** y T **Wilusz**, *J Mol Biol* (1989), 210:635-648.

HOLA89b:

Holak, TA, W **Bode**, R **Huber**, J **Otlewski** y T **Wilusz**, *J Mol Biol* (5 de diciembre de 1989), 210(3)649-54.

HORV89:

Horvat, S, B **Grgas**, N **Raos** y VI **Simeon**, *Int J Peptide Protein Res* (1989), 34:346-51.

HOUG84:

Houghten, RA, JM **Ostresh** y FA **Klipstein**, *Eur J Biochem* (1984), 145:157-162.

ITOK79:

Ito, K, G **Mandel** y W **Wickner**, *Proc Natl Acad Sci USA* (1979), 76:1199-1203.

JANA89:

Janatova, J, KBM **Reid** y AC **Willis**, *Biochem* (1989), 28:4754-61.

JANI85:

Janin, J y C **Chothia**, *Methods in Enzymology* (1985), 115(28)420-430.

JENN89:

Jennings, PA, MM **Bills**, DO **Irving** y JS **Mattick**, *Protein Eng*, (enero de 1989), 2(5)365-9.

JUDD85:

Judd, RC, *Infect Immun* (1985), 48(2)452-7.

JUDD86:

Judd, RC, *Infect Immun* (1986), 54(2)408-14.

KATZ86:

Katz, BA y A **Kossiakoff**, *J Biol Chem* (1986), 261(33)15480-85.

KATZ90:

Katz, B y AA **Kossiakoff**, *Proteins, Struct, Funct, and Genet* (1990), 7:343-57.

KISH85:

Kishore, R y P **Balaram**, *Biopolymers* (1985), 24:2041-43.

KOBA89:

Kobayashi, Y, T **Ohkubo**, Y **Kyogoku**, Y **Nishiuchi**, S **Sakakibara**, W **Braun** y N **Go**, *Biochemistry* (1989), 28:4853-60.

KUB089:

Kubota, H, Y **Hidaka**, H **Ozaki**, H **Ito**, T **Hirayama**, Y **Takeda** y Y **Shimonishi**, *Biochem Biophys Res Comm* (1989), 161:229-235.

KUHN85a:

Kuhn, A y W **Wickner**, *J Biol Chem* (1985), 260:15914-15918.

KUHN85b:

Kuhn, A y W **Wickner**, *J Biol Chem* (1985), 260:15907-15913.

KUPE90:

Kupersztoch, YM, K **Tachias**, CR **Moomaw**, LA **Dreyfus**, R **Urban**, C **Slaughter** y S **Whipp**, *J Bacteriol* (1990), 172(5)2427-32.

LAM91:

Lam, KS.*et al*. *Nature* (1991), 354:82-84.

LISS85:

Liss, LR, BL **Johnson** y DB **Oliver**, *J Bacteriol* (1985), 164(2)925-8.

LUND86:

Lundeen, M, *J Inorgan Biochem* (1986), 27:151-62.

MACI88:

Maclntyre, S, R **Freudl**, ML **Eschbach** y U **Henning**, *J Biol Chem* (1988), 263(35)19053-9.

MANI82:

Maniatis, T, EF **Fritsch** y J **Sambrook**, *Molecular Cloning, Cold Spring Harbor Laboratory*, 1982.

MAN086:

Manoil, C y J **Beckwith**, *Science* (1986), 233:1403-1408.

MAN088:

Manoil, C, D **Boyd** y J **Beckwith**, *Topics in Genetics* (1988), 4(8)223-6.

MARK86:

Marks, CB, M **Vasser**, P **Ng**, W **Henzel** y S **Anderson**, *J Biol Chem* (1986), 261:7115-7118.

MARK91:

Markland *et al*. *Gene* (1991) 109:13-19.

MATS89:

Matsumura, M, WJ **Becktel**, M **Levitt** y BW **Matthews**, *Proc Natl Acad Sci USA* (1989), 86:6562-6.

MCCA90:

McCafferty, J, AD **Griffiths**, G **Winter** y DJ **Chiswell**, *Nature*, (6 de diciembre de 1990), 348:552-4.

MCKE85:

McKern, NM, IJ **O'Donnell**, DJ **Stewart** y BL **Clark**, *J Gen Microbiol* (1985), 131(Pt 1)1-6.

MCWH89:

McWherter, CA, WF **Walkenhorst**, EJ **Campbell** y Gl **Glover**, *Biochemistry* (1989), 28:5708-14.

MISR88a:

Misra, R y SA **Benson**, *J Bacteriol* (1988), 170(8)3611-7.

MISR88b:

Misra, R y SA **Benson**, *J Bacteriol* (1988), 170:528-33.

MORS87:

Morse, SA, TA **Mietzner**, G **Bolen**, A Le **Faou** y G **Schooinik**, *Antonie Van Leeuwenhoek* (1987), 53(6) 465-9.

MORS88:

Morse, SA, C-Y **Chen**, A **LeFaou** y TA **Meitzner**, *Rev Infect Dis* (1988), 10(Supl. 2)S306-10.

NICH88:

Nicholson, H, WJ **Becktel** y BW **Matthews**, *Nature* (1988), 336:651-56.

NIKA84:

Nikaido, H y HCP **Wu**, *Proc Natl Acad Sci USA* (1984), 81:1048-52.

NISH82:

Nishiuchi, Y y S **Sakakibara**, *FEBS Lett* (1982), 148:260-2.

NISH86:

Nishiuchi, Y, K **Kumagaye**, Y **Noda**, TX **Watanabe** y S **Sakakibara**, *Biopolymers*, (1986), 25:S61-8.

OKAM87:

Okamoto, K, K **Okamoto**, J **Yukitake**, Y **Kawamoto** y A **Miyama**, *Infection and Immunity* (1987), 55:2121-2125.

OKAM88:

Okamoto, K, K **Okamoto**, J **Yukitake** y A **Miyama**, *Infection and Immunity* (1988), 56:2144-8.

OKAM90:

Okamoto, K y M **Takahara**, *J Bacterid* (1990), 172(9)5260-65.

OLIP86:

Oliphant, AR, AL **Nussbaum** y K **Struhl**, *Gene* (1986), 44:177-183.

OLIP87:

Oliphant, AR y K **Struhl** *Methods in Enzymoloqy 155* (1987)568-582. Editor Wu, R; Academic Press, New York.

OLIV87:

Olivera, BM, LJ **Cruz**, V de **Santos**, GW **LeCheminant**, D **Griffin**, R **Zeikus**, JM **Mcintosh**, R **Galyean**, J **Varga**, WR **Gray**, *et al*. *Biochemistry*, (1987), 26(8)2086-90.

OLIV90a:

Olivera, BM, J **Rivier**, C **Clark**, CA **Ramilo**, GP **Corpuz**, FC **Abogadie**, EE **Mena**, SR **Woodward**, DR **Hillyard**, LJ **Cruz**, *Science*, (20 de julio de 1990), 249:257-263.

ORND85:

Orndorff, PE y S **Falkow**, *J Bacteriol* (1985), 162:454-7.

OTLE87:

Otlewski, J, H **Whatley**, A **Polanowski** y T **Wilusz**, *Biol Chem Hoppe-Seyler* (1987), 368:1505-7.

PAB079:

Pabo, CO, RT **Sauer**, JM **Sturtevant** y M **Ptashne**, *Proc Natl Acad Sci USA* (1979), 76:1608-1612.

PAB086:

Pabo, CO y EG **Suchanek**, *Biochem* (1986), 25:5987-91.

PAGE88:

Pages, JM y JM **Bolla**, *Eur J Biochem* (1988), 176(3)655-60.

PAGE90:

Pages, JM, JM **Bolla**, A **Bemadac** y D **Fourel**, *Biochimie* (1990), 72:169-76.

PAKU86:

Pakula, AA, VB **Young** y RT **Sauer**, *Proc Natl Acad Sci USA* (1986), 83:8829-8833.

PANT87:

Pantoliano, MW, RC **Ladner**, PN **Bryan**, ML **Rollence**, JF **Wood** y TL **Poulos**, *Biochem* (1987), 26:2077-82.

PANT90:

Pantoliano, MW y RC **Ladner**, Patente de Estados Unidos Nº 4.908.773, 13 de marzo de 1990.

PARD89:

Pardi, A, A **Galdes**, J **Florance** y D **Maniconte**, *Biochemistry* (1989), 28:5494-5501.

PARG87:

Parge, HE, DE **McRee**, MA **Capozza**, SL **Bernstein**, ED **Getzoff** y JA **Tainer**, *Antonie Van Leeuwenhoek* (1987), 53(6)447-53.

PARM88:

Parmley, SF y GP **Smith**, *Gene* (1988), 73:305-318.

PARR88:

Parraga, G, SJ **Horvath**, A **Eisen**, WE **Taylor**, L **Hood**, ET **Young**, RE **Klevit**, *Science* (1988), 241:1489-92.

PEAS88:

Pease, JHB y DE **Wemmer**, *Biochem* (1988), 27:8491-99.

PEAS90:

Pease, JHB, RW **Storrs** y DE **Wemmer**, *Proc Natl Acad Sci USA* (1990), 87:5643-47.

PERR84:

Perry, LJ y R **Wetzel**, *Science* (1984), 226:555-7.

PERR86:

Perry, LJ y R **Wetzel**, *Biochem* (1986), 25:733-39.

POTE83:

Poteete, AR, *J Mol Biol* (1983), 171:401-418.

QUI087:

Quiocho, FA, NK **Vyas**, JS. **Sack** y MA **Storey**, en *Crystallography in Molecular Biology*, Moras, D. *et al*, editores, *Plenum Press*, 1987.

RASC86:

Rasched, I y E **Oberer**, *Microbiol Rev* (1986) 50:401-427.

RASH84:

Rashin, A, *Biochemistry* (1984), 23:5518.

REID88a:

Reidhaar-Olson, JF y RT **Sauer**, *Science* (1988), 241:53-57.

RTSD88b:

Reid, J, H **Fung**, K **Gehring**, PE **Klebba** y H **Nikaido**, *J Biol Chem* (1988), 263(16)7753-9.

RICH86:

Richards, JH, *Nature* (1986), 323:187.

ROBE86:

Roberts, S y AR **Rees** *Protein Engineering* (1986), 1:59-65.

R0SS81:

Rossman, M y P **Argos**, *Ann Rev Biochem* (1981), 50:497-532.

SALI88:

Sali, D, M **Bycroft** y AR **Fersht**, *Nature* (1988), 335:740-3.

SAUE86:

Sauer, RT, K **Hehir**, RS **Stearman**, MA **Weiss**, A **Jeitler-Nilsson**, EG **Suchanek** y CO **Pabo**, *Biochem* (1986), 25:5992-98.

SCHA78:

Schaller, H, E **Beck** y M **Takanami**, The Single-Stranded DNA Phages, Denhardt, D.T., D. Dressier y D.S. Ray editors, *Cold Spring Harbor Laboratory*, 1978., págs. 139-163.

SCHN86:

Schnabel, E, W **Schroeder** y G **Reinhardt**, *Biol Chem Hoppe-Seyier* (1986), 367:1167-76.

SCHU79:

Schulz, GE y RH **Schirmer**, Principles of Protein Structure, *Springer-Verlag*, New York, 1979.

SCOT87a:

Scott, MJ, CS **Huckaby**, I **Kato**, WJ **Kohr**, M **Laskowski** Jr., M-J **Tsai** y BW **O'Malley**. *J Biol Chem* (1987), 262(12)5899-5907.

SCOT90:

Scott, JK y GP **Smith**, *Science*, (27 de julio de 1990), 249:386-390.

SEKI85:

Sekizaki, T, H **Akaski** y N **Terakado**, *Am 1 Vet Res* (1985), 46:909-12.

SHIM87:

Shimonishi, Y, Y **Hidaka**, M **Koizumi**, M **Hane**, S **Aimoto**, T **Takeda**, T **Miwatani** y Y **Takeda**, *FEBS Lett* (1987), 215:165-170.

SILH77:

Silhavy, TJ, HA **Shuman**, J **Beckwith** y M **Schwartz**, *Proc Natl Acad Sci USA* (1977), 74(12)5411-5415.

SMIT85:

Smith GP, *Science* (1985), 228:1315-1317.

SODE85:

Sodergren, EJ, J **Davidson**, RK **Taylor** y TJ **Silhavy**, *J Bacteriol* (1985), 162(3)1047-1053.

SOME85:

So, M, E **Billyard**, C **Deal**, E **Getzoff**, P **Hagblom**, TF **Meyer**, E **Segal** y J **Tainer**, *CurrTop in Microbiol & Immunol* (1985), 118.13-28.

STAD89:

Stader, J, LJ **Gansheroff** y TJ **Silhavy**, *Genes & Develop* (1989), 3:1045-1052.

STEI85:

Steiner, *BioScience Repts*. (1985), 5:973ff.

SUMM91:

Summers, MF, *J Cell Biochem* (1991) 45:41-8.

SUNX87:

Sun, XP, H **Takeuchi**, Y **Okano** y Y **Nozawa**, *Comp Biochem Physiol* [C], (1987), 87(2)363-6.

TAKA85:

Takao, T, N **Tominaga**, S **Yoshimura**, Y **Shimonishi**, S **Hara**, T **Inoue** y A **Miyama**, *Eur J Biochem* (1985), 152:199-206.

TAKE90:

Takeda, T, GB **Nair**, K **Suzuki** y Y **Shimonishi**, *Infection and Immunity* (1990), 58(9)2755-9.

TANK77:

Tan, NH y ET **Kaiser**, *Biochemistry*, (1977), 16:1531-41.

THOM85a:

Thompson, MR, M **Luttrell**, G **Overmann**, RA **Giannella**, *Analytical Biochem* (1985), 148:26-36.

THOM85b:

Thompson, MR y RA **Giannella**. *Infection & Immunity* (1985), 47:834-36.

THOR88:

Thornton, JM, BL **Sibinda**, MS **Edwards** y DJ **Barlow**, *BioEssays* (?) SKG 3039 ??????

TOMM82:

Tommassen, J, P van der **Ley**, A van der **Ende**, H **Bergmans** y B **Lugtenberg**, *Mol gen Genet* (1982), 185:105-110.

TRIA88:

Trias, J, EY **Rosenberg** y H **Nikaido**, *Biochim Biophys Acta* (1988), 938:493-496.

VAND86:

van der **Ley**, P, M **Struyve** y J **Tommassen**, *J Biol Chem* (1986), 261(26)12222-5.

VAND90:

van der **Werf**, S, A **Charbit**, C **Leclerc**, V **Mimic**, J **Ronco**, M **Girard** y M **Hofnung**, *Vaccine* (1990), 8(3)269-77.

VERS86a:

Vershon, AK, K **Blacker** y RT **Sauer**, págs. 243-256 en *Protein Engineering. Applications in Science. Medicine, and Industry*, *Academic Press*, 1986.

VERS86b:

Vershon, AK, JU **Bowie**, TM **Karplus** y RT **Sauer**, págs. 302-311 en *Proteins: Structure. Function, and Genetics, Alan R. Liss, Inc.*, 1986.

VITA84:

Vita, C, D **Dalzoppo** y A **Fontana**, *Biochemistry* (1984), 23:5512-5519.

VOGE86:

Vogel, H y F **Jahnig**, *J Mol Biol* (1986), 190:191-99.

WATS87:

Molecular Biology of the Gene, Fourth Edition **Watson**, JD, NH **Hopkins**, JW **Roberts**, JA **Steitz** y AM **Weiner**, Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA., 1987.

WEBS78:

Webster, RE y JS **Cashman**, *The Single-Stranded DNA Phages*, Denhardt, DT, D Dressier y DS Ray editors, *Cold Spring Harbor Laboratory*, 1978., págs. 557-569.

WEHM89:

Wehmeier, U, GA **Sprenger** y JW **Lengeler**, *Mol Gen Genet* (1989), 215(3)529-36.

WEIN83:

Weinstock, GM, C ap **Rhys**, ML **Berman**, B **Hampar**, D **Jackson**, TJ **Silhavy**, J **Weisemann** y M **Zweig**, *Proc Natl Acad Sci USA* (1983), 80:4432-4436.

WELL86:

Wells, JA y DB **Powers**, *J Biol Chem* (1986); 261:6564-70.

WEMM83:

Wemmer, D y NR **Kallenbach**, *Biochem* (1983), 22:1901-6.

WHAR86:

Wharton, RP, *The Binding Specificity Determinants of 434 Repressor.,* Harvard U. PhD Thesis, 1986, University Microfilms, Ann Arbor, Michigan.

WIEC85:

Wieczorek, M, J **Otlewski**, J **Cook**, K **Parks**, J **Leluk**, A **Wilimowska-Pelc**, A **Polanowski**, T **Wilusz** y L **Laskowski**, Jr, *Biochem Biophys Res Comm* (1985), 126(2)646-652.

WILK84:

Wilkinson, AJ, AR **Fersht**, DM **Blow**, P **Carter** y G **Winter**, *Nature* (1984), 307:187-188.

WOOD90:

Woodward, SR, IJ **Cruz**, BM **Olivera** y DR **Hillyard**, *EMBO J* (1990), 9:1015-1020.

YOSH85:

Yoshimura, S, H **Ikemura**, H **Watanabe**. S **Aimoto**, Y **Shimonishi**, S **Hara**,T **Takeda**, T **Miwatani** y Y **Takeda**, *FEBS Lett* (1985), 181:138-42.

ZAFA88:

Zafaralla, GC, C **Ramilo**, WR **Gray**, R **Karlstrom**, BM **Olivera** y LJ **Cruz**, *Biochemistry*, (1988), 27(18) 7102-5.

REIVINDICACIONES

1. Una biblioteca de proteínas quiméricas, comprendiendo cada proteína quimérica:

a) una micro-proteína de entre 6 y 40 aminoácidos que tiene dos o más puentes disulfuro, donde cada puente disulfuro está formado por un par de cisteínas invariantes, y

b) una secuencia de aminoácidos obtenida a partir de una proteína de la superficie externa de un paquete genético, en la que la proteína quimérica se presenta en la superficie externa del paquete genético.

2. La biblioteca de la reivindicación 1, en la que dos o más cisteínas están agrupadas.

3. La biblioteca de la reivindicación 1, en la que cada tramo del enlace disulfuro es mayor de 9 aminoácidos.

4. La biblioteca de la reivindicación 1, en la que la micro-proteína tiene dos puentes disulfuro.

5. La biblioteca de la reivindicación 1, en la que la proteína tiene tres puentes disulfuro.

6. La biblioteca de una cualquiera de las reivindicaciones 1 a 5, en la que los aminoácidos de la secuencia de la micro-proteína, diferentes de cisteína, son variables.

7. La biblioteca de una cualquiera de las reivindicaciones 1 a 5, en la que un tramo de los puentes disulfuro comprende aminoácidos variables.

8. La biblioteca de las reivindicaciones 1-7, en la que un puente disulfuro abarca una secuencia de aminoácidos que colectivamente adopta una estructura secundaria en forma de horquilla.

9. La biblioteca de la reivindicación 8, en la que la estructura secundaria en forma de horquilla es una hélice alfa, un giro y una cadena beta.

10. La biblioteca de la reivindicación 8, en la que la estructura secundaria en forma de horquilla es una hélice alfa, un giro y una hélice alfa.

11. La biblioteca de la reivindicación 8, en la que la estructura secundaria en forma de horquilla es una cadena beta, un giro y una cadena beta..

12. La biblioteca de una cualquiera de las reivindicaciones 9-11, en la que no más de tres aminoácidos están situados entre una cisteína invariante de un par de cisteinas invariantes y el comienzo de la hélice alfa o cadena beta más próxima.

13. La biblioteca de una cualquiera de las reivindicaciones 9-11, en la que no más de tres aminoácidos están situados entre una cisteína invariante de un par de cisteínas invariantes y el final de la hélice alfa o cadena beta más próxima.

14. La biblioteca de una cualquiera de las reivindicaciones 9 a 11, en la que los aminoácidos variables dentro de una estructura secundaria de hélice alfa o cadena beta están limitados a posiciones para alterar de manera poco probable la estructura secundaria, determinandose o diseñándose la estructura en base al análisis estructural de la estructura actual a partir de una proteína y/o datos de frecuencia.

15. La biblioteca de una cualquiera de las reivindicaciones 6-14, en la que al menos un aminoácido que se ha modificado en la micro-proteína está codificado por un codón variegado seleccionado entre NNT, NNG, RNG, RMG, VNT, RRS, SNT.

16. La biblioteca de una cualquiera de las reivindicaciones 6-15, en la que ninguno de los aminoácidos que se ha modificado en la micro-proteína está codificado por un codón variegado seleccionado entre NNN, NNK y NNS.

17. La biblioteca de una cualquiera de las reivindicaciones 1-16, en la que la microproteína está directamente fusionada a al menos una parte de la proteína de la superficie externa.

18. La biblioteca de una cualquiera de las reivindicaciones 1-17, en la que el paquete genético es un fago filamentoso.

19. La biblioteca de una cualquiera de las reivindicaciones 1-18, en la que la proteína de la superficie externa es la proteína de cubierta principal de un fago filamentoso.

20. La biblioteca de una cualquiera de las reivindicaciones 1-19, en la que la proteína de la superficie externa es la proteína del gen III de un fago filamentoso.

21. La biblioteca de una cualquiera de las reivindicaciones 1-17, en la el paquete génico es una célula bacteriana.

22. La biblioteca de la reivindicación 21, en la que la célula bacteriana es *Salmonella typhimurium, Bacilus subtilis, Pseudomonas aeruginosa, Vibrio cholerae, Klebsiella pneumonia, Neisseria gonorrhoeae, Neisseria meningitidis, Bacteroides nodosus, Moraxella bobis* o *Escherichia coli.*

23 La biblioteca de una cualquiera de las reivindicaciones 1-22, en la que la micro-proteína se une a una diana.

24. La biblioteca de la reivindicación 23, en la que la diana es una serina proteasa.

25. La biblioteca de la reivindicación 24, en la que la serina proteasa es elastasa neutrófila humana.

26. Un proceso para identificar proteínas con una actividad de unión deseada contra una diana que comprende:

a) explorar una biblioteca de proteínas quiméricas de acuerdo con una cualquiera de las reivindicaciones 1-25; y

b) identificar la proteína quimérica.

27. Una proteína quimérica expresada por una biblioteca de acuerdo con una cualquiera de las reivindicaciones 1-25.

28. Una mezcla de ácidos nucleicos, donde la mezcla codifica la biblioteca de una cualquiera de las reivindicaciones 1-25.

29. La mezcla de la reivindicación 28, en la que cada ácido nucleico se clona en un plásmido, un virus, un fagémido o un cromosoma.

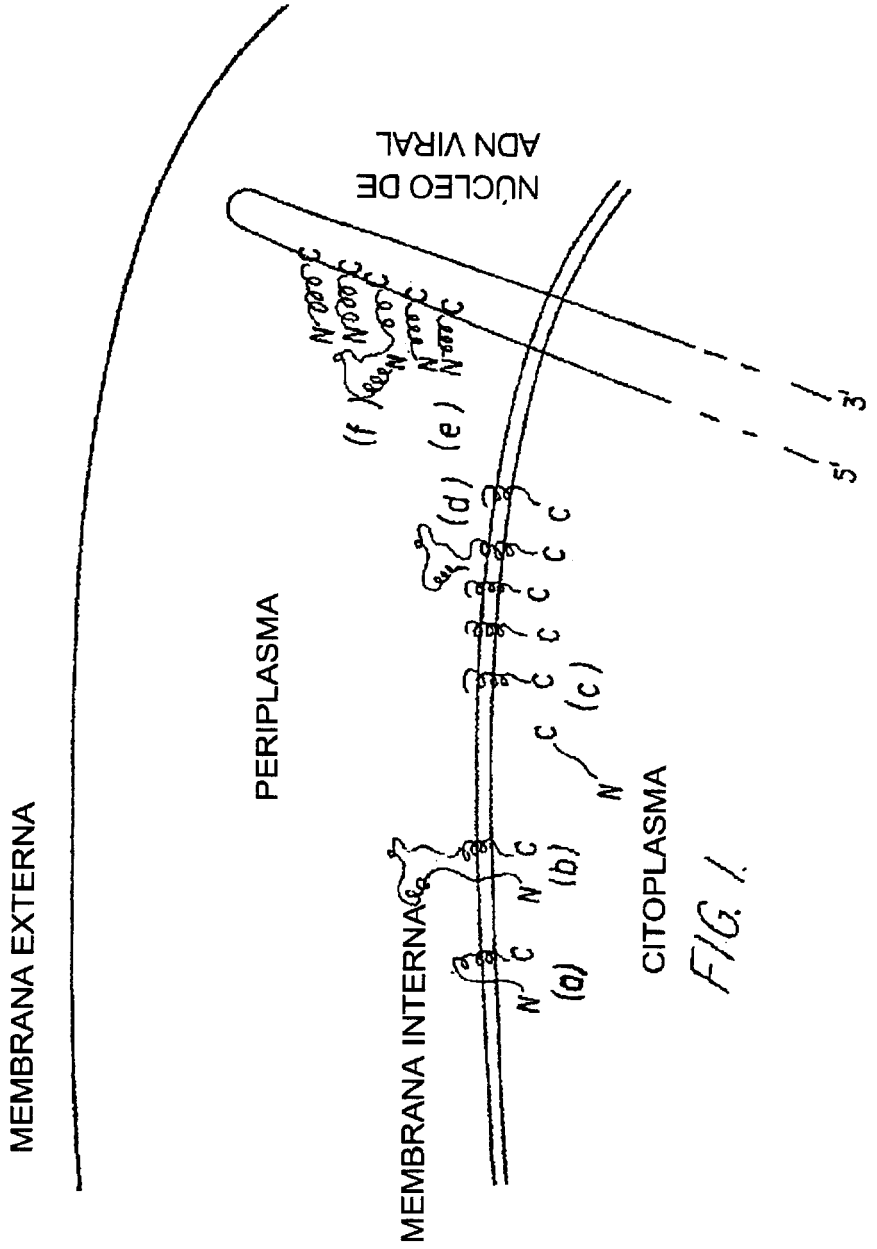

MEMBRANA EXTERNA
PERIPLASMA
MEMBRANA INTERNA
CITOPLASMA
NÚCLEO DE
ADN VIRAL
5'
3'
(a)
(b)
(c)
(d)
(e)
(f)
N
C

FIG. 1.