

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 547 121**

51 Int. Cl.:

C07K 1/00 (2006.01)

C07K 7/08 (2006.01)

C07K 14/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **08.12.2009 E 09774880 (0)**

97 Fecha y número de publicación de la concesión europea: **12.08.2015 EP 2367840**

54 Título: **Proteínas de superhélice antiparalela de una sola cadena**

30 Prioridad:

08.12.2008 US 120642 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

01.10.2015

73 Titular/es:

**COMPLIX N.V. (100.0%)
Agoralaan Building A-bis
3590 Diepenbeek, BE**

72 Inventor/es:

**DESMET, JOHAN;
LASTERS, IGNACE y
LOVERIX, STEFAN**

74 Agente/Representante:

FÚSTER OLAGUIBEL, Gustavo Nicolás

Observaciones :

Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 547 121 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Proteínas de superhélice antiparalela de una sola cadena

5 **Campo de la invención**

La presente invención se encuentra en el campo de la biología molecular y se refiere a proteínas de una sola cadena, termodinámicamente estables, que consisten esencialmente en una estructura de armazón de superhélice alfa-helicoidal, antiparalela, de triple hebra en disoluciones acuosas. Tales moléculas son muy estables y tolerantes a sustituciones de aminoácidos. Por consiguiente, cumplen los requisitos básicos de un armazón basado en proteínas. Este armazón que muestra capacidades terapéuticas, de diagnóstico y/o de purificación, puede usarse en el campo del descubrimiento de fármacos, la investigación analítica, la tecnología de purificación y como modelo para mejorar el diseño de nuevas estructuras de armazón proteínicas (similares a proteínas). Las moléculas de armazón basadas en proteínas a menudo se consideran la clase de compuestos de "nueva generación" para el reconocimiento molecular, que compiten cada vez más con compuestos basados en inmunoglobulinas. Por consiguiente, los compuestos de la presente invención ofrecen un enfoque alternativo a las inmunoglobulinas, y un tipo adicional de armazón basado en proteínas (proteínico).

20 **Antecedentes de la invención**

Los complejos de superhélice alfa-helicoidal de triple hebra (3 hebras) (estructuras de superhélice, superhélices) se forman en disolución por la asociación (reunión) de moléculas peptídicas individuales (separadas, monoméricas, libres) dando lugar a trímeros (complejos de 3 moléculas). Los péptidos individuales comprenden normalmente una o más repeticiones en héptada (unidades de héptada, héptadas) que proporcionan la fuerza motriz termodinámica para tal asociación.

Un problema práctico importante encontrado con la formación de complejos triméricos es el hecho de que tales reacciones son extremadamente dependientes de la concentración. Por tanto, a menos que la fuerza motriz electrodinámica sea extremadamente fuerte (es decir, sólo si las héptadas forman interacciones extremadamente estrechas), tienen que aplicarse concentraciones relativamente altas con el fin de que se forme el complejo trimérico. Concentraciones altas pueden tener múltiples efectos adversos cuando se aplican a (se administran como) compuestos farmacéuticos. A diferencia de los complejos triméricos, la formación de (el plegamiento de) estructuras de superhélice de una sola cadena de la presente invención no es dependiente de su concentración en disolución. La presente invención por tanto pretende proporcionar una solución al problema de la dependencia de la concentración.

Un segundo problema relacionado con el uso de complejos oligoméricos peptídicos (multiméricos) es que los péptidos constituyentes son difíciles de producir (sintetizar) a través de métodos recombinantes (es decir, usando técnicas biológicas moleculares). Esto contrasta con las proteínas de una sola cadena plegadas de manera estable, que son adecuadas idealmente para la síntesis recombinante. Por tanto, la presente invención proporciona una solución a los problemas técnicos relacionados con la síntesis de armazones de superhélice triméricos en forma peptídica.

En tercer lugar, la presente invención tiene como objetivo proporcionar una solución práctica al problema de crear estructuras de superhélice heterotriméricas. La naturaleza oligomérica de las superhélices peptídicas se define en general por el número de péptidos asociados (por ejemplo, 2, 3, 4, para complejos diméricos, triméricos, tetraméricos, respectivamente), su orientación mutua (por ejemplo, paralela o antiparalela) y su similitud química (es decir, su secuencia de aminoácidos con derivatización opcional; por ejemplo, las superhélices homotriméricas se forman mediante tres péptidos idénticos, las superhélices heterotriméricas comprenden al menos un péptido derivatizado o de secuencia diferente). Las superhélices oligoméricas pueden obtenerse en disolución acuosa mezclando péptidos no idénticos. Entonces, tras un tiempo de incubación suficientemente largo, se formará una distribución de superhélices homo y heteroméricas, dependiendo principalmente de la capacidad termodinámica de estas últimas (estabilidad, energía libre, calidad de la asociación). En vista de las interacciones atómicas complicadas que constituyen la base de la capacidad termodinámica y, de ese modo, de las preferencias (distribuciones) oligoméricas, la creación de tipos específicos, deseados de superhélices heteroméricas es técnicamente difícil de controlar. A este respecto, la presente invención proporciona una solución práctica a un problema técnico: puesto que los fragmentos peptídicos que forman la superhélice están unidos covalentemente entre sí dando lugar a una sola cadena (a través de fragmentos de ligador elegidos de manera adecuada), su propensión a formar estructuras de superhélice de naturaleza predefinida (deseada) está considerablemente potenciada en comparación con superhélices equivalentes que consisten en conjuntos de péptidos libres. Por consiguiente, la construcción de superhélices heterodiméricas específicas (por ejemplo, heterotriméricas) se facilita considerablemente. Además, el formato de superhélice de una sola cadena también ofrece la ventaja de evitar (o reducir considerablemente el riesgo de) la formación de tipos de asociación no deseados (por ejemplo, no funcionales). En general, el formato de una sola cadena, que se aplica a todas las realizaciones de la presente invención, proporciona una solución práctica para controlar y conservar la especificidad de plegamiento de una superhélice trimérica en la que los fragmentos peptídicos que forman la superhélice son (opcionalmente) diferentes

en la secuencia de aminoácidos.

Todas las realizaciones de la presente invención se refieren a estructuras de superhélice alfa-helicoidal de “una sola cadena” aunque de “triple hebra”. Por motivos de claridad, se explica en el presente documento (y se comenta adicionalmente más adelante en detalle) que la propiedad de “una sola cadena” se refiere a las moléculas completas de la presente invención, mientras que la propiedad de “triple hebra” se refiere a la parte de superhélice alfa-helicoidal dentro de estas moléculas. Siempre que se use la descripción “superhélice de una sola cadena”, esto debe interpretarse como una asociación estrecha entre (tres) fragmentos peptídicos que forman superhélice que están interconectados covalentemente mediante (dos) fragmentos de ligador estructuralmente flexibles; dichos fragmentos peptídicos y de ligador forman juntos una molécula de proteína que consiste en una sola cadena de aminoácidos, contigua. Las proteínas de superhélice de una sola cadena de la presente invención también son monómeros (moléculas de proteína monoméricas en disolución), lo que no debe confundirse con la naturaleza trimérica de la estructura de superhélice que está contenida dentro de cada una de tales proteínas. La proteína de una sola cadena aislada se define en las reivindicaciones.

La inmensa mayoría de las estructuras de superhélice de triple hebra en el Protein Data Bank (Banco de Datos de Proteínas) (a continuación en el presente documento denominado PDB) son superhélices paralelas, es decir del tipo de “péptidos alfa-helicoidales paralelos”. Esto significa que las superhélices existen como complejos (asociaciones no covalentes) de tres péptidos alfa-helicoidales por estructura y en las que las hélices están orientadas en una configuración (orientación) paralela. Muy raramente, una de las tres alfa-hélices se orienta antiparalela a las otras dos (que entonces son paralelas entre sí). Una disposición antiparalela de este tipo es excepcional en las proteínas naturales y nunca se ha observado en forma de una estructura de superhélice regular que se compone de, y se estabiliza mediante, motivos de repetición en héptada convencionales. La tabla 1 muestra una lista exhaustiva de 179 complejos peptídicos de superhélice de triple hebra a partir del PDB, 175 de los cuales son paralelos y sólo 4 son antiparalelos. Esto sugiere que una orientación paralela es la configuración más estable para las superhélices triméricas peptídicas. Un motivo probable para la abundancia de configuraciones paralelas es la conservación de simetría de 3 pliegues, que permite un número máximo de contactos óptimos. En cambio, todas las realizaciones de la presente invención se refieren a superhélices de una sola cadena que adoptan una orientación antiparalela. En vista de los escasos ejemplos de estructuras de superhélice de triple hebra antiparalela en el PDB, el diseño y la creación de tales estructuras no son obvios en absoluto. Por ejemplo, un trabajo de este tipo no solo es complicado por la falta de estructuras de molde representativas (ejemplo), tampoco está claro a priori si pueden desarrollarse superhélices antiparalelas con interacciones de núcleo de calidad comparable tal como se observa en superhélices de triple hebra paralelas. En vista de lo anterior, uno de los principales aspectos inventivos de la presente invención es el hallazgo inesperado de que pueden obtenerse superhélices de triple hebra antiparalelas altamente estables. Esto indica que los residuos de núcleo en posiciones de repetición en héptada convencionales también pueden realizar interacciones casi óptimas en una configuración antiparalela, lo que se desconocía anteriormente.

Sumario de la invención

Los inventores han construido estructuras proteicas de superhélice de triple hebra de una sola cadena que se esperaba que se plegaran en configuración paralela, pero con fragmentos de ligador que eran significativamente demasiado cortos como para permitir este tipo de plegamiento. Inesperadamente, se encontró que estos últimos constructos tenían las mismas propiedades físicas (contenido alfa-helicoidal, estabilidad térmica, solubilidad, etc.) que variantes con ligadores muy largos. Mientras que, en general, los constructos con ligadores físicamente demasiado cortos provocan el desplegamiento de la estructura, las estructuras de armazón triméricas de la presente invención presentaban inesperadamente alta estabilidad térmica en condiciones que se desviaban significativamente de las condiciones fisiológicas, por ejemplo en urea 8 M, o a temperaturas que superaban los 90°C, y esto independientemente de las longitudes del ligador. Estos hallazgos sugieren fuertemente que las moléculas de la presente invención se pliegan dando lugar a una configuración antiparalela. Esto último también se confirmó mediante espectroscopía de RMN. Tales estructuras de superhélice nuevas por consiguiente son de alto valor para muchas aplicaciones basadas en armazón.

La presente invención se refiere a una clase de proteínas de una sola cadena, nuevas, aisladas, de fórmula HRS1-L1-HRS2-L2-HRS3, en la que HRS1, L1, HRS2, L2 y HRS3 representan fragmentos de secuencia de aminoácidos que están interconectados covalentemente, plegándose espontáneamente dicha proteína en disolución acuosa por medio de los fragmentos HRS1, HRS2 y HRS3 que forman una estructura de superhélice alfa-helicoidal, antiparalela, de triple hebra, y en la que

cada uno de HRS1, HRS2 y HRS3 es independientemente una secuencia de repetición en héptada que se caracteriza por un patrón de 7 residuos repetido n veces de tipos de aminoácidos, representado como (a-b-c-d-e-f-g)-n o (d-e-f-g-a-b-c)-n y una repetición en héptada parcial C-terminal que termina con un residuo de núcleo de héptada ubicado en una posición a o d, en la que los elementos de patrón “a” a “g” indican posiciones de héptada convencionales en las que se ubican dichos tipos de aminoácidos y n es un número igual a o mayor de 2, y

al menos el 50%, el 70%, el 90% o el 100% de las posiciones de héptada convencionales “a” y “d” están ocupadas por aminoácidos seleccionados del grupo que consiste en valina, isoleucina, leucina, metionina, fenilalanina, tirosina,

triptófano, histidina, glutamina, treonina, serina, alanina o derivados de los mismos, en la que al menos el 50%, el 70%, el 90% o en la que el 100% de las posiciones de héptada convencionales “b”, “c”, “e”, “f” y “g” están ocupadas por aminoácidos seleccionados del grupo que consiste en glicina, alanina, cisteína, serina, treonina, histidina, asparagina, ácido aspártico, glutamina, ácido glutámico, lisina, arginina o derivados de los mismos, permitiendo la distribución resultante de los tipos de aminoácidos la identificación de dichas secuencias de repetición en héptada, y

cada uno de L1 y L2 es independientemente un ligador que consiste en de 1 a 30 residuos de aminoácido, incluyendo este ligador cualquier residuo de aminoácido que no pueda asignarse de manera inequívoca a una secuencia de repetición en héptada, y

L1 y L2 tienen una composición de aminoácidos que comprende al menos el 50% de aminoácidos seleccionados del grupo que consiste en glicina, alanina, serina, treonina, prolina o derivados de los mismos.

Breve descripción de los dibujos

La figura 1 ilustra una secuencia de aminoácidos de un péptido sintético que comprende repeticiones en héptada. La secuencia de aminoácidos se presenta en una notación de una sola letra, en la que A se refiere a alanina, I a isoleucina, Q a glutamina y K a lisina. El péptido comprende repeticiones en héptada (HRx), residuos de núcleo (recuadros negros), residuos no de núcleo (recuadros grises) y regiones flanqueantes (recuadros blancos). El péptido comprende además un residuo de núcleo de héptada C-terminal marcado como “t”. El péptido comprende además fragmentos flanqueantes N- y C-terminales marcados como “N” y “C”, respectivamente. Cada residuo de repetición en héptada se anota además con índices “a” a “g” y un número correspondiente al número de repetición en héptada. Los residuos de núcleo se ubican en las posiciones a y d.

La figura 2 ilustra los principios de un complejo de superhélice alfa-helicoidal, de triple hebra. La figura proporciona una representación en rueda helicoidal de estructuras de superhélice alfa-helicoidal, de triple hebra. El panel de la izquierda muestra una vista desde arriba de una superhélice paralela. El panel de la derecha muestra una vista desde arriba de una superhélice antiparalela. El panel del medio muestra la secuencia lineal de posiciones de repeticiones en héptada. Solo se muestra una repetición en héptada por motivos de claridad. Se usan diferentes tonos para indicar posiciones topológicas específicas.

La figura 3 ilustra la desnaturalización térmica de una superhélice peptídica, monitorizada mediante dicroísmo circular (DC). Se muestra el espectro de DC del péptido Ac-MSIEEIQQQAAIQKQIAAIQKQIYRMTN-NH₂ a 5 y 90 grados Celsius (curvas negra y gris, respectivamente). El péptido se disolvió a una concentración de 292 microM en tampón fosfato (PBS) 20 mM, NaCl 150 mM, pH 7,2.

La figura 4 ilustra el plegamiento y desplegamiento reversible del péptido de la figura 3, monitorizado mediante la señal de DC a 222 nM en función de la temperatura (se muestran barridos ascendentes y descendentes).

La figura 5 ilustra el análisis termodinámico adicional de la curva de desplegamiento térmico de la figura 4. La curva negra representa datos experimentales tomados de la figura 4, mientras que la curva blanca representa la curva ajustada. La curva teórica (ajustada) se obtuvo mediante el procedimiento explicado en el ejemplo 3. Los parámetros ajustados (resultados de ajuste) se enumeran a la derecha de la figura 5. “T. de transición” corresponde a T_t , pero se expresa en grados Celsius. El parámetro “delta C_p ” se mantuvo constante a $3,0 \text{ kJ mol}^{-1} \text{ K}^{-1}$. Los parámetros “theta_M (T)” y “theta_T (T)” se trataron como funciones lineales de T, dando como resultado las líneas rectas discontinuas descritas por las respectivas desviaciones y pendientes indicadas a la derecha de la figura. “RMS Resid.” se refiere a la raíz cuadrática media de las diferencias entre puntos de datos experimentales y teóricos.

La figura 6 ilustra la curva de barrido térmico de DC para una preparación de muestra del péptido Q2a1 en las mismas condiciones que en el ejemplo 3. El péptido Q2a1 tiene la secuencia de aminoácidos Ac-MSIEEIQQQIAAIQKQIAAIQKQIYRMTN-NH₂. Los resultados de un barrido ascendente y descendente se muestran en negro y gris, respectivamente.

La figura 7 ilustra los resultados de ultracentrifugación en equilibrio de sedimentación analítica para el péptido Q2a1 de la figura 6. La curva de sedimentación se obtuvo a 25000 rotaciones por minuto (rpm). La figura muestra la curva de densidad óptica (DO) linealizada en comparación con las curvas teóricas para complejos monoméricos, diméricos y triméricos, tal como se indica mediante los marcadores.

La figura 8 ilustra los resultados de dispersión de luz estática para el péptido Q2a1 de la figura 6. Se pusieron 200 microlitros de péptido a 1 mg/ml en PBS en una columna de filtración en gel Superdex 75 10/300 GL conectada a detectores de ultravioleta (UV), de índice de refracción (RI) y de dispersión de luz estática (SLS). Las señales (curvas) de los tres detectores diferentes se marcan en consecuencia.

La figura 9 ilustra las secuencias de aminoácidos de dos proteínas que forman realizaciones específicas de la presente invención. Estas dos proteínas se denominan “scQ2a1_L8” y “scQ2a1_L16”, respectivamente. Sus secuencias de aminoácidos completas se enumeran al final de cada panel de tabla, a la derecha de la marca

“Completa”. También se muestran segmentos específicos dentro de las mismas secuencias en la parte superior, para facilitar identificación de segmentos flanqueantes N- y C-terminales (marcados como “N” y “C”, respectivamente), segmentos de ligador (marcados como “L1” y “L2”, respectivamente) y la secuencia real de repetición en héptada (marcada como “HRS1”, “HRS2” y “HRS3”). Las posiciones a y d de héptada se facilitan en la fila superior para facilitar su identificación dentro de las secuencias de repetición en héptada.

La figura 10 ilustra el barrido térmico de DC para el constructo scQ2al_L16. Se registró el barrido para este constructo en PBS 20 mM, NaCl 150 mM, pH 7,2.

La figura 11 ilustra la desnaturalización térmica de scQ2al_L16 y scQ2al_L8 (marcados en consecuencia) en GuHCl 6 M registrada mediante DC a 222 nm en tampón PBS y a una concentración de proteína de aproximadamente 30 μ M. Se ajustaron los barridos térmicos a un modelo de transición de dos estados y se convirtieron en fracción de proteína plegada.

La figura 12 ilustra las temperaturas de transición de diversos constructos que forman realizaciones específicas de la presente invención, en función de la concentración de GuHCl (desnaturalizante). Dichos constructos se denominan “scQ2al_L16”, “short_L6”, “short_L10”, “short_L14” y “short_L18”, y las curvas correspondientes se marcan en consecuencia. Las secuencias de dichos constructos, un método para producirlos y las condiciones experimentales se detallan adicionalmente en el ejemplo 5.

La figura 13 muestra los espectros de RMN de ^{15}N ^1H HSQC para los constructos scQ2al_L16 y scQ2al_L8 (marcados en consecuencia).

La figura 14 muestra una ampliación en el espectro de RMN de un mutante doble de triptófano-cisteína con marcador de espín del constructo scQ2al_L16, tal como se explica en el ejemplo 6. Se registró el espectro en la muestra no tratada y en una muestra tratada con vitamina C (resonancias marcadas en consecuencia).

La figura 15 muestra modelos moleculares de superhélices de una sola cadena de 3 hebras antiparalelas (marcadas en consecuencia). Se prepararon los modelos tal como se explica en el ejemplo 7. Las tres alfa-hélices en cada modelo se marcan como “A”, “B” y “C” y representan secuencias de repetición en héptada HRS1, HRS2 y HRS3 en dichas superhélices de una sola cadena, respectivamente. Los marcadores “L1” y “L2” indican los segmentos de ligador respectivos. “Nt” y “Ct” indican los extremos N- y C-terminales de cada constructo, respectivamente.

Descripción detallada de la invención

El término “armazón” se usa dentro del contexto de la presente invención para indicar “una molécula proteínica (similar a proteína o proteína) específica, conformacional (estructural) y termodinámicamente (térmica y químicamente) estable con una estructura (disposición espacial de elementos constituyentes) tridimensional (3-D, terciaria) específica, fija (invariable, invariante) que consiste en una o más cadenas proteicas o polipeptídicas proteínicas, siendo tolerante dicha estructura de manera demostrable a una variedad de sustituciones de aminoácidos individuales y múltiples en una variedad de posiciones de residuos de aminoácidos.

La noción “tolerante a sustituciones de aminoácidos” ha de entenderse en el presente documento en el sentido de que la integridad (corrección) de la estructura permanece esencialmente inalterada tras realizarse dichas sustituciones de aminoácidos. Es evidente que cualquier sustitución de aminoácido en una proteína altera la estructura 3-D en cierto grado, pero tales cambios son del dominio público y en el presente documento se consideran no esenciales si la estructura principal de la proteína (cadena principal) de la estructura 3-D mutada (sustituida) sigue pudiendo superponerse estructuralmente con la estructura no mutada (original, de tipo natural); se considera que dos estructuras pueden superponerse si al menos el 70% de los átomos de la estructura principal (excluyendo los átomos de hidrógeno) de ambas estructuras pueden superponerse con una desviación de raíz cuadrática media (RMS) de preferiblemente menos de 1 Ångström (1 Å), menos preferiblemente de 2 Å o 3 Å. En casos en los que no es factible una superposición estructural (por ejemplo si no está disponible una de las dos estructuras 3-D), entonces la noción “tolerante a sustituciones de aminoácidos” ha de interpretarse en el sentido termodinámico: se considera que una proteína es tolerante a sustitución/sustituciones de aminoácidos si la(s) sustitución/sustituciones disminuyen el punto medio de la transición térmica (temperatura de transición, T_t , temperatura de fusión, T_m , temperatura de desplegamiento T_u) preferiblemente en no más de 10 grados Celsius ($^{\circ}\text{C}$) en comparación con el tipo natural, menos preferiblemente en no más de 20 $^{\circ}\text{C}$, o 30 $^{\circ}\text{C}$, o 40 $^{\circ}\text{C}$, o 50 $^{\circ}\text{C}$, y en cualquier caso no hasta el punto de que la proteína sustituida se despliegue cuantitativamente a temperatura fisiológica (37 $^{\circ}\text{C}$). La propiedad “tolerante a una variedad de sustituciones en una variedad de posiciones” pretende significar en el presente documento tolerante a al menos aproximadamente 10 residuos de aminoácido diferentes en al menos 5 posiciones de aminoácido diferentes, más preferiblemente en 10 posiciones, o 20 posiciones, lo más preferiblemente en aproximadamente el 50% o más de todas las posiciones de aminoácido.

La esencia de lo que se entiende generalmente por una molécula de armazón es una molécula que actúa como portador de grupos químicos. De manera similar, las proteínas de armazón (o, brevemente, armazones) en el presente documento se refieren a proteínas o moléculas proteínicas que sirven como portadores de cadenas

laterales de aminoácidos. También pueden servir como portadores de otras proteínas, o fragmentos, dominios o péptidos que se unen a cualquiera de sus extremos terminales (es decir, como parte de un constructo de fusión), pero éste no es el significado previsto dentro del presente contexto. Puesto que las cadenas laterales de aminoácidos en una proteína se unen a la cadena principal (estructura principal), la estructura principal plegada constituye formalmente la forma químicamente más pura de un almacén. Sin embargo, las estructuras principales proteicas puras, con poli-glicina como análogo de polipéptido más próximo, no se pliegan de manera estable en disolución, y por tanto no cumplen los requisitos de un almacén útil. Por consiguiente, las proteínas que se privan parcial o completamente de sus cadenas laterales no forman el objeto de la presente invención. En cambio, la presente invención reivindica proteínas reales que adoptan un plegamiento 3-D dado (en este caso, una estructura de superhélice alfa-helicoidal, antiparalela, de triple hebra, de una sola cadena) y que hacen esto de una manera termodinámicamente estable, incluso tras haber experimentado un número sustancial de mutaciones. Por tanto, el término “almacén” se refiere a su robustez estructural y termodinámica, más que a una función portadora.

Las moléculas de proteína de la presente invención pueden usarse como almacenes, de manera similar a muchos otros almacenes documentados (revisado en Skerra [J Mol Recognit 2000, 13:167-187], Binz *et al.* [Nat Biotechnol 2005, 23:1257-1268], Hosse *et al.* [Protein Sci 2006, 15:14-27]). La noción “usada como almacén” significa esencialmente que las moléculas deseadas (por ejemplo, con una determinada funcionalidad) pueden obtenerse (derivarse) a partir de un constructo de referencia preseleccionado (almacén de referencia). Las moléculas derivadas son normalmente variantes con sustitución de aminoácido o sustitución de bucle del almacén de referencia.

Las moléculas de almacén basadas en proteínas (proteínicas) distintas de inmunoglobulinas se consideran en el campo como una clase de compuestos de “próxima generación” para el reconocimiento molecular. Se derivan principalmente de moléculas de proteína naturales que se han seleccionado basándose en propiedades fisicoquímicas preferidas y datos experimentales disponibles. Ejemplos de esta clase de compuestos se enumeran por Hosse *et al.* [Protein Sci 2006, 15:14-27] y por Binz *et al.* [Nat Biotechnol 2005, 23:1257-1268].

La presente invención divulga un tipo particular de moléculas de proteína distinta de inmunoglobulina que tienen excelentes propiedades para su uso como almacenes de proteína. Debido a su alta estabilidad y robustez estructural, pueden construirse grandes bibliotecas (bibliotecas basadas en almacenes, bibliotecas de almacenes) de moléculas con esencialmente las mismas estructuras terciarias y secuencias ligeramente diferentes. Alternativamente, pueden variarse los residuos de superficie haciendo uso de métodos convencionales de ingeniería de proteínas. Haciendo uso del conocimiento del experto, pueden aplicarse métodos de selección apropiados con el propósito de identificar variantes (derivados de almacén, compuestos moleculares específicos) con propiedades de unión altamente deseadas (por ejemplo, afinidades y especificidades) similares a inmunoglobulinas.

Se han atribuido numerosas ventajas a las moléculas de almacén basadas en proteínas con respecto a las inmunoglobulinas incluyendo, por ejemplo, su tamaño relativamente pequeño, alta estabilidad estructural y ausencia de modificaciones postraduccionales. Estas características facilitan considerablemente su síntesis, purificación y almacenamiento. Además, pueden generarse compuestos de alta afinidad sin necesidad de proceder a través de una etapa de inmunización. Los almacenes de proteína de la presente invención incorporan todas las características mencionadas anteriormente, lo que las convierte en particularmente muy adecuadas para aplicaciones basadas en almacén.

La presente invención se refiere a un tipo particular de almacén basado en proteínas que es en gran medida insensible a la sustitución de residuos de superficie y a las acciones convencionales de ingeniería de proteínas. Todas las realizaciones de la presente invención se refieren a un tipo específico de estructura de proteínas (estructura 3-D, estructura terciaria, plegamiento) que hasta ahora no se ha explotado como almacén de proteína altamente mutable, en este caso una estructura de superhélice alfa-helicoidal, antiparalela, de triple hebra, de una sola cadena.

Las proteínas de la presente invención tienen un amplio espectro de posibles aplicaciones, en gran medida comparables a las de las inmunoglobulinas. Más concretamente, los mutantes derivados de almacén específicos pueden usarse como compuestos terapéuticos (por ejemplo, inhibidores), sondas de detección (por ejemplo, detección de una proteína recombinante) y sondas de purificación (por ejemplo, en cromatografía de afinidad), tal como se detalla a continuación en el presente documento.

Las moléculas de proteína de la presente invención pueden ser adecuadas como compuestos terapéuticos. Más específicamente, pueden interferir (influir, modificar) con procesos biológicos a través de impedir (bloquear, inhibir) reacciones químicas naturales o acontecimientos de reconocimiento molecular naturales, o a través de la creación de acontecimientos de reconocimiento molecular no naturales. Los ejemplos de interferencia biológica incluyen, sin limitación, el bloqueo de receptores humanos, la unión a especies patógenas y la unión a proteínas relacionadas con enfermedades o trastornos. Tal tipo de interferencia biológica está destinada normalmente a curar enfermedades o trastornos graves. Estas aplicaciones pertenecen al campo de la investigación y el desarrollo terapéuticos. Los tratamientos terapéuticos actuales se basan generalmente en compuestos farmacológicos o biotecnológicos, incluyendo estos últimos o bien compuestos (derivados) de inmunoglobulina o bien compuestos distintos de

inmunoglobulina. La producción, purificación, pruebas y optimización de ambos tipos de compuestos biotecnológicos generalmente requiere mucho trabajo, es arriesgado y caro. Por consiguiente, existe la necesidad de nuevos compuestos biotecnológicos con actividad biológica específica, así como métodos mejorados para la producción, purificación, pruebas y optimización de tales compuestos.

5 Las moléculas de proteína de la presente invención pueden ser adecuadas como sondas de detección. Los ejemplos en los que se aplican moléculas de sonda (sondas) específicas para detectar la presencia de un analito de interés (analito diana) en una muestra de interés (muestra de estudio) dada, incluyen, sin limitación, análisis experimentales de muestras de origen humano, animal, vegetal, bacteriano, viral, biotecnológico o sintético. Tales muestras
10 contienen normalmente biomoléculas (por ejemplo, polipéptidos, polinucleótidos, polisacáridos, hormonas, vitaminas o lípidos, o derivados de los mismos) que pueden interactuar específicamente con una molécula de sonda seleccionada. Esta última interacción normalmente da lugar a una señal (por ejemplo, espectroscópica o radiactiva) característica, indicativa de la presencia de dicho analito diana en dicha muestra de estudio. Estas aplicaciones pertenecen al campo de la investigación y el desarrollo analíticos. El número de combinaciones de diferentes tipos de sondas y dianas que se usan eficazmente en aplicaciones médicas y biotecnológicas es prácticamente ilimitado.
15 En vista de la continua evolución en estas áreas, hay una necesidad en curso de nuevas herramientas analíticas (por ejemplo, sondas) con propiedades fisicoquímicas deseadas (por ejemplo, especificidad, afinidad, estabilidad, solubilidad), así como métodos mejorados para la producción, purificación, pruebas y optimización de tales compuestos.

20 Las moléculas de proteína de la presente invención pueden ser adecuadas para aplicaciones de purificación. Los ejemplos en los que se aplican moléculas de ligando (ligandos) específicas para retener (extraer, aislar, purificar, filtrar) otras moléculas de interés (dianas, analitos diana) en una muestra de interés (muestra en bruto) dada incluyen, sin limitación, muestras de origen humano, animal, vegetal, bacteriano, viral, biotecnológico o sintético que
25 contienen biomoléculas (por ejemplo, polipéptidos, polinucleótidos, polisacáridos, hormonas, vitaminas o lípidos, o derivados de los mismos) que pueden interactuar (asociarse) con alta especificidad con moléculas de ligando seleccionadas, donde estas últimas se separan, o pueden separarse, de la muestra en bruto (por ejemplo, mediante la unión sobre un soporte sólido o mediante precipitación), con el propósito de separar conjuntamente las moléculas diana de la muestra en bruto. Estas aplicaciones pertenecen al campo de la tecnología de purificación. Los ejemplos
30 más específicos de métodos de purificación incluyen cromatografía de afinidad e inmunoprecipitación. En vista de la continua evolución en estas áreas, hay una necesidad en curso de nuevos ligandos para purificación con propiedades fisicoquímicas deseadas (por ejemplo, especificidad, afinidad, estabilidad, solubilidad), así como métodos mejorados para la producción, purificación, pruebas y optimización de tales compuestos.

35 Las moléculas de armazón de proteína de la presente invención se pliegan dando lugar a una estructura de superhélice alfa-helicoidal. La superhélice alfa-helicoidal forma un tipo especial de marco estructural (motivo estructural, plegamiento) 3-D. El plegamiento de superhélice se produce en una amplia variedad de proteínas incluyendo proteínas motoras, proteínas de unión a ADN, proteínas extracelulares y proteínas de fusión viral (por ejemplo, Burkhard *et al.* [Trends Cell Biol 2001, 11:82-88]). Se ha estimado que del 3 al 5%, o más, de todos los
40 aminoácidos en las proteínas naturales forman parte de una estructura de superhélice [Wolf *et al.*, Protein Sci 1997, 6:1179-1189].

45 Las superhélices se han caracterizado funcionalmente como motivos de plegamiento (ensamblaje, oligomerización), es decir, la formación de una estructura de superhélice conduce en muchos casos a la asociación no covalente de diferentes cadenas de proteína. Las superhélices se han caracterizado estructuralmente como conjuntos de 2, 3, 4 ó 5 hebras de alfa-hélices dispuestas en topologías paralelas, antiparalelas o mixtas (por ejemplo, Lupas [Trends Biochem Sci 1996, 21:375-382]. Las hélices están ligeramente empaquetadas (enroscadas, enrolladas) entre sí hacia la izquierda o hacia la derecha, denominados superenrollamiento. Todas las realizaciones de la presente invención se refieren exclusivamente a estructuras de superhélice de triple hebra (3 hebras, triméricas).

50 Las superhélices alfa-helicoidales se han caracterizado adicionalmente a nivel de sus secuencias de aminoácidos, en que cada hélice está constituida por una serie de repeticiones en héptada. Una repetición en héptada (unidad de héptada, héptada) es un motivo de secuencia de 7 residuos que puede codificarse como HppHppp, y en la que cada "H" representa un residuo hidrófobo (potencialmente diferente) y cada "p" es un residuo polar (potencialmente diferente). De manera ocasional (infrecuentemente), se observan residuos p en posiciones de H, y viceversa. Una
55 repetición en héptada de la invención se codifica por los patrones a-b-c-d-e-f-g (abcdefg) o d-e-f-g-a-b-c (defgabc), en cuyo caso los índices "a" a "g" se refieren a las posiciones de héptada convencionales en las que se observan tipos de aminoácidos típicos. Por convenio, los índices "a" y "d" indican las posiciones de los residuos de núcleo (residuos centrales, escondidos) en una superhélice. Los tipos de aminoácidos típicos que se observan en las
60 posiciones de núcleo a y d son tipos de residuos de aminoácidos hidrófobos; en el resto de las posiciones (posiciones no de núcleo), se observan tipos de residuos predominantemente polares (hidrófilos). Por tanto, los patrones de héptada convencionales "HppHppp" coinciden con la notación de patrón "abcdefg" (los patrones "HpppHpp" coinciden con la notación de patrón "defgabc", usándose esta notación para superhélices que comienzan con un residuo hidrófobo en la posición d). Todas las realizaciones de la presente invención incluyen al menos 2,
65 preferiblemente 3 o más repeticiones en héptada consecutivas (ininterrumpidas) en cada alfa-hélice de la estructura de superhélice. Cada serie de repeticiones en héptada consecutivas en una hélice se indica como una "secuencia de

repetición en héptada" (HRS). El comienzo y el final de una secuencia de repetición en héptada se determinan preferiblemente basándose en la estructura tridimensional (3-D) determinada experimentalmente, si está disponible. Si no se dispone de una estructura 3-D, el comienzo y el final de una secuencia de repetición en héptada se determina preferiblemente basándose en una superposición óptima de un patrón $(\text{HppHppp})_n$ o $(\text{HpppHpp})_n$ con la secuencia de aminoácidos real, donde "H" y "p" indican residuos hidrófobos y polares, respectivamente, y donde "n" es un número igual a o mayor de 2. Entonces, se considera que el comienzo y el final de cada secuencia de repetición en héptada es el primer y el último residuo hidrófobo en una posición a o d, respectivamente. Los residuos H convencionales se seleccionan preferiblemente del grupo que consiste en valina, isoleucina, leucina, metionina, fenilalanina, tirosina, triptófano, histidina, glutamina, treonina, serina y alanina, más preferiblemente del grupo que consiste en valina, isoleucina, leucina y metionina, y lo más preferiblemente isoleucina. Los residuos p convencionales se seleccionan preferiblemente del grupo que consiste en glicina, alanina, cisteína, serina, treonina, histidina, asparagina, ácido aspártico, glutamina, ácido glutámico, lisina y arginina. En caso de que este método sencillo no permita la asignación inequívoca de residuos de aminoácido a una secuencia de repetición en héptada, puede aplicarse un método de análisis más especializado, tal como el método COILS de Lupas *et al.* [Science 1991, 252:1162-1164; <http://www.russell.embl-heidelberg.de/cgi-bin/coils-svr.pl>].

Las superhélices se han caracterizado termodinámicamente de la siguiente forma. Cuando la secuencia se pliega dando lugar a una alfa-hélice, los residuos hidrófobos (H) forman una unión hidrófoba, mientras que los residuos polares (p) forman una cara polar. Las uniones hidrófobas de diferentes alfa-hélices, cuando se asocian dando lugar a una superhélice, forman un núcleo hidrófobo central (centro, interior, parte interna). Se supone que la formación de este núcleo, en combinación con la orientación de las caras polares hacia el disolvente, proporciona la principal fuerza motriz termodinámica requerida para la asociación estable, aunque determinados residuos no de núcleo también pueden potenciar la estabilidad. Todas las realizaciones de la presente invención se refieren a estructuras de superhélice de triple hebra que consisten en al menos dos repeticiones en héptada por alfa-hélice y en las que los residuos H de las repeticiones en héptada forman el núcleo hidrófobo y, como tales, proporcionan la principal fuerza motriz termodinámica para el plegamiento de la estructura.

Las superhélices de triple hebra peptídicas (no de una sola cadena) pueden mostrar una alta estabilidad térmica pese a su dependencia de la oligomerización y, por tanto, alta dependencia de la concentración. Por ejemplo, se mostró que la cremallera de Ile de Suzuki *et al.* [Protein Eng 1998, 11:1051-1055] tenía una temperatura de fusión (desplegamiento, transición) que superaba los 80°C. De manera similar, Harbury *et al.* [Science 1993, 262:1401-1407; Nature 1994, 371:80-83] diseñaron una superhélice de triple hebra derivada de GCN4, denominada GCN4-pII, que se encontró que era estable en forma cristalina y en disolución. Además, también se diseñaron con éxito superhélices paralelas heterotriméricas [Nautiyal y Alber, Protein Sci 1999, 8:84-90]. También se conocen a grandes rasgos las principales normas para que los péptidos se unan dando lugar a configuraciones paralelas triméricas [Yu, Adv Drug Deliv Rev 2002, 54:1113-1129]. Además, la solicitud internacional PCT/EP2008/061886 ha reivindicado superhélices de 3 hebras peptídicas en forma de un amazón proteínico, termodinámicamente estable no natural. Las moléculas de la presente invención también comprenden una estructura de superhélice de 3 hebras, pero difieren fundamentalmente de las superhélices peptídicas (que forman complejos triméricos) en que están compuestas por una sola cadena de aminoácido que se pliega como una proteína monomérica.

Aunque lo anterior puede sugerir que el diseño de superhélices paralelas de 3 hebras es relativamente sencillo, muchos estudios han notificado graves dificultades. Por ejemplo, una superhélice que se diseñó como un dímero paralelo se observó en la estructura cristalina como un trímero antiparalelo [Lovejoy *et al.*, Science 1993, 259:1288-1293]. Además, el requisito de una secuencia desencadenante para potenciar la cinética de plegamiento ha sido objeto de debate [Yu, *ibídem*]. Además, el proceso de desplegado térmico no siempre sigue un mecanismo de dos estados simple [Dragan y Privalov, J mol Biol 2002, 321:891-908] y el proceso de ensamblaje (plegamiento) ocasionalmente es muy lento [Dragan *et al.*, Biochemistry 2004, 43:14891-14900]. Por consiguiente, en vista de los muchos resultados inesperados obtenidos por investigadores expertos pese a la abundancia de datos experimentales sobre superhélices paralelas, puede concluirse que el diseño y la aplicación de moléculas de superhélice alfa-helicoidal incluso paralela son absolutamente no obvios. Por consiguiente, el desarrollo de superhélices antiparalelas puede preverse todavía más complicado.

Los inventores contemplaron inicialmente el uso de amazones de superhélice de triple hélice peptídicos, mientras que al mismo tiempo se intentaba encontrar una solución práctica a las desventajas inherentes de tales complejos, que tienen que trimerizarse primero en disolución antes de adoptar el plegamiento apropiado (es decir, deseado, funcional). Tal solución se encontró finalmente en forma de una versión de una sola cadena de un trímero, en la que el extremo C-terminal (C-terminal) de una primera alfa-hélice constituyente se conecta (se une, se asocia) al extremo N-terminal (N-terminal) de una segunda alfa-hélice, y el extremo C-terminal de esta última al extremo N-terminal de una tercera alfa-hélice. Según la terminología de Harris *et al.* [J Mol Biol 1994, 236:1356-1368], las conexiones entre hélices paralelas se denominan conexiones "hacia abajo" (o "largas"), y entre hélices antiparalelas se denominan conexiones "hacia arriba" (o "cortas"). En las realizaciones de la presente invención, las conexiones entre alfa-hélices consecutivas se realizan a través del uso de fragmentos de ligador estructuralmente flexibles, dando lugar a constructos en los que se unen entre sí tres alfa-hélices mediante dos ligadores flexibles. Todas las realizaciones de la presente invención pertenecen a este tipo de disposición. Las moléculas de la presente invención pueden escribirse formalmente por tanto como una secuencia de fórmula HRS1-L1-HRS2-L2-HRS3, en la que HRS1, L1,

HRS2, L2 y HRS3 representan fragmentos de secuencia de aminoácidos que están interconectados de manera covalente y consecutiva en el orden indicado en dicha fórmula, y en la que los fragmentos HRS1, HRS2 y HRS3 son secuencias de repetición en héptada tal como se describió anteriormente, y en la que los fragmentos L1 y L2 son secuencias de ligador estructuralmente flexibles.

Los fragmentos de ligador flexibles se usan frecuentemente en el campo de la ingeniería de proteínas para interconectar diferentes unidades funcionales, por ejemplo en la creación de constructos de fragmento variable de cadena sencilla (scFv) derivados de cadenas variable ligera (VL) y variable pesada (VH) de anticuerpos. En la actualidad, la aplicación de fragmentos de ligador flexibles en combinación con estructuras de superhélice triméricas, con el propósito de crear una estructura de armazón de superhélice de una sola cadena aunque de triple hebra no se ha divulgado ni se prevé en el dominio público. También se observa que no hay ninguna contradicción en la formulación “de una sola cadena aunque de triple hebra” porque “una sola cadena” se refiere a la secuencia de aminoácidos completa, mientras que “triple hebra” es el término común para indicar que la estructura de superhélice consiste en tres hebras (fragmentos de cadena) alfa-helicoidales individuales. Todas las realizaciones de la presente invención comprenden exactamente dos segmentos (fragmentos) de ligador flexibles dentro del contexto de una estructura de superhélice de 3 hebras. Los segmentos de ligador no son necesariamente idénticos en longitud ni en secuencia de aminoácidos. Sin embargo, para potenciar la probabilidad de que sean conformacionalmente flexibles en disolución, están compuestos de manera preferible y predominante por tipos de residuos aminoácidos polares. Los aminoácidos típicos (frecuentemente usados) en ligadores flexibles son serina y glicina. Menos preferiblemente, los ligadores flexibles también pueden incluir alanina, treonina y prolina. Todavía menos preferidos (debido al aumento del riesgo de interacciones no deseadas) es la incorporación de cisteína, histidina, asparagina, ácido aspártico, glutamina, ácido glutámico, lisina y arginina, o derivados no naturales de los mismos, en combinación con dichos aminoácidos más preferidos.

Un método preferido y sencillo para distinguir los fragmentos de ligador de las secuencias de repetición en héptada es determinar el primer lugar esta última mediante cualquiera de los métodos descritos anteriormente, y luego incluir los fragmentos de aminoácidos restantes en los ligadores. Este método se aplica tanto al caso en el que no existe estructura 3-D determinada experimentalmente de la molécula de proteína como al caso en el que sí existe una o más de tales estructuras. Si tal(es) estructura(s) determinada(s) experimentalmente diera(n) lugar a incertidumbre o ambigüedad con respecto al estado estructuralmente flexible de cualquiera de los ligadores, entonces la noción “ligador flexible” debe interpretarse simplemente como un fragmento que puede conectar (unir, formar un puente) entre dos secuencias de repetición en héptada, y no como un fragmento estructuralmente dinámico o móvil.

El uso de ligadores flexibles en la presente invención pretende principalmente interconectar los fragmentos alfa-helicoidales con el propósito de crear una secuencia de aminoácidos lineal (constructo de una sola cadena). Aunque esto es técnicamente sencillo, un aspecto importante que ha de considerarse es la longitud (el número de residuos de aminoácido) de cada ligador. Para las superhélices paralelas en las que las hélices comprenden el mismo número de residuos, la distancia en el espacio 3-D desde el extremo (C-terminal) de una alfa-hélice dada hasta el inicio (N-terminal) de una alfa-hélice adyacente (conexión hacia abajo) puede calcularse de manera aproximada mediante la fórmula “número de residuos por alfa-hélice, multiplicado por 1,5 Ångström”. La distancia en que un ligador puede formar un puente en conformación extendida puede calcularse aproximadamente mediante la fórmula “número de residuos en el fragmento de ligador, multiplicado por 3,0 Ångstrom”. Por tanto, como norma, un ligador debe tener al menos la mitad del número de residuos por alfa-hélice para permitir la conexión hacia abajo de manera relajada. (Se aplican excepciones a esta norma cuando las hélices son de diferente longitud o cuando los giros de hélice con respecto a ligador no se realizan fácilmente: en tales casos, preferiblemente se añade un número pequeño de residuos de ligador adicionales).

De manera importante, dicha norma proporciona una forma práctica de calcular la longitud de ligador mínima necesaria para una conexión hacia abajo entre elementos alfa-helicoidales en configuración paralela, y no un método para imponer una orientación paralela. La conformación de un ligador flexible en disolución será, o al menos se pretende que sea esencialmente al azar en estructura y dinámica en comportamiento (es decir, estructuralmente variable en el tiempo). Por tanto, un ligador “de longitud suficiente” permitirá, pero no impondrá, el plegamiento en paralelo. A la inversa, un ligador “de longitud insuficiente” (“ligador demasiado corto”) no permitirá el plegamiento en paralelo y por tanto inducirá o bien desplegamiento o bien formación de un plegamiento alternativo (siempre que este último sea estable por sí mismo). Una posibilidad de este tipo de un plegamiento alternativo es una estructura de superhélice antiparalela: los requisitos para la unión entre hélices antiparalelas (conexión hacia arriba) son topológicamente muy complejos, pero son generalmente menos restrictivos. En otras palabras, un ligador que es significativamente demasiado corto como para formar un puente en la distancia entre las alfa-hélices en orientación paralela puede permitir muy bien el plegamiento antiparalelo. De manera importante, esto último no implica que tal ligador corto se requiere o será necesario para inducir el plegamiento antiparalelo (esto último depende esencialmente de la posibilidad de la formación de un núcleo física y termodinámicamente estable en modo antiparalelo, posiblemente potenciado además por interacciones favorables adicionales entre residuos no de núcleo).

Puesto que se ha observado que las estructuras de superhélice triméricas se pliegan, con raras excepciones, en orientación paralela, es poco probable que las mismas secuencias puedan adoptar también un plegamiento

antiparalelo estable. Esto último es de relevancia específica para la presente invención, porque los inventores han generado y caracterizado estructuras de superhélice de triple hebra de una sola cadena que estaban dotadas de ligadores que eran significativamente demasiado cortos para el plegamiento paralelo, mientras que la molécula todavía se plegaba con conservación completa del contenido alfa-helicoidal y con efectos insignificantes sobre la temperatura de transición en experimentos de desplegamiento térmico (véase el ejemplo 5). Basándose en estos experimentos, se concluyó que estos constructos (y posiblemente también los que tienen ligadores largos) adoptan presumiblemente un plegamiento antiparalelo.

Para someter a prueba si el plegamiento antiparalelo es estructuralmente factible, los inventores han intentado generar modelos 3-D de una superhélice trimérica de una sola cadena en los que la segunda alfa-hélice ("B") es antiparalela con respecto a la primera ("A") y la tercera ("C"). Inesperadamente, pudieron generarse modelos creíbles con empaquetamiento de "botones en ojales" regular mediante operaciones de modelado de proteínas convencionales (véase el ejemplo 7). Todas las cadenas laterales de formación de núcleo pudieron colocarse en su conformación rotamérica más relajada. Resulta interesante que las posiciones a de héptada convencionales de la hélice B antiparalela se empaquetan sobre los residuos d de las hélices A y C (capas d). De esta forma, la hélice B interacciona con A y C a lo largo de toda su longitud, lo que sugiere que todas las posiciones de núcleo de héptada contribuyen a la estabilidad del plegamiento. Aunque esto no demuestra que se trata de plegamiento antiparalelo, los resultados del modelado sugieren que es al menos estructuralmente factible, a diferencia de las suposiciones originales.

Lovejoy *et al.* [Science 1993, 259:1288-1293] realizaron una observación inesperada similar para "Coil-Ser", un péptido que se diseñó para formar una superhélice paralela de doble hebra, pero realmente ensamblado dando lugar a una superhélice de triple hebra. Esta estructura se estabilizó mediante una superficie de contacto hidrófoba, distintiva, no deliberada que consiste en ocho capas (se encontró que cada capa a dentro de las hélices paralelas estaba asociada con un residuo d de la hélice antiparalela, y cada capa d estaba asociada con un residuo a; las capas se denominaron "a-a-d" y "d-d-a", respectivamente). En otro estudio realizado por Holton y Alber [Proc Natl Acad Sci USA 2004, 101:1537-1542], un mutante de Ala en la cremallera de leucina de GCN4 también cambió desde la configuración de dímero paralela por defecto a una configuración de trímero antiparalela. Se encontró que este cambio estructural se debía a que se evitaba crear cavidades en el núcleo. Se encontró la misma disposición en capas a-a-d y d-d-a alternas como en el estudio de Holton y Alber. De manera importante, ambos estudios se referían a superhélices que tenían un núcleo formado por residuos de leucina ("cremalleras de Leu"), mientras que los presentes inventores observaron una orientación antiparalela para superhélices que tenían un núcleo formado por residuos de isoleucina ("cremalleras de Ile"); nunca antes se había encontrado que las cremalleras de Ile formaran superhélices de 3 hebras antiparalelas. En segundo lugar, ambos de dichos estudios se referían a superhélice peptídicas, mientras que las moléculas de la presente invención se refieren exclusivamente a superhélices de una sola cadena; nunca antes se habían obtenido superhélices de 3 hebras antiparalelas en forma de moléculas de una sola cadena (formato de una sola cadena). No se sabe si la orientación antiparalela de estas moléculas de proteína se debe a la presencia de los fragmentos de ligador, o a sus secuencias de aminoácidos específicas, o a cualquier otro motivo, o a una combinación de motivos. En cualquier caso, el modelo 3-D de las moléculas de la presente invención, así como las estructuras cristalográficas descritas en los citados estudios de Lovejoy *et al.* y Holton y Alber [ibídem], son todas ellas estructuras de superhélice verdaderas con residuos de núcleo empaquetados de manera regular en capas separadas de manera regular. Esto las distingue de haces de tres hélices ordinarios.

Las estructuras de superhélice antiparalela de triple hélice no deben confundirse con haces de tres haces ordinarios: hay muchos ejemplos de asociaciones (haces) de tres alfa-hélices que no son superhélices regulares. Pueden observarse haces de alfa-hélices en un gran número de proteínas principalmente alfa-helicoidales, y a menudo pueden discernirse haces de tres hélices que interaccionan mutuamente dentro de tales proteínas. Las superhélices de triple hebra evidentemente también son haces de tres alfa-hélices, pero para que un haz de 3 hélices sea una superhélice es necesario que se cumplan varias condiciones adicionales. En primer lugar, se requiere que las tres hélices interaccionen mutuamente entre sí, lo que excluye topologías en las que sólo dos de los tres pares posibles de hélices están en contacto entre sí (topologías no cohesivas). En segundo lugar, debe haber un grado apropiado de superenrollamiento (es decir, empaquetamiento de las hélices unas alrededor de las otras). El principal determinante de superenrollamiento es el ángulo entre cada par de hélices (ángulo interhelicoidal, ángulo de interacción hélice-hélice, ángulo de cruzamiento). Para alfa-hélices paralelas, este ángulo puede variar desde pequeños valores negativos para superenrollamiento "dextrógiro", normalmente en el intervalo de aproximadamente -10 grados a 0 grados, hasta valores positivos para superenrollamiento "levógiro", normalmente en el intervalo de aproximadamente 20 grados a 0 grados. Para alfa-hélices antiparalelas, deben restarse 180 grados de dichos valores. Las topologías con un ángulo demasiado grande, fijado este último a 40 grados en valor absoluto, no se consideran superhélices. En tercer lugar, debe haber repeticiones en héptada discernibles dentro de cada una de las alfa-hélices que interaccionan (tal como se definió anteriormente). Las verdaderas superhélices comprenden al menos 2, preferiblemente al menos 3 repeticiones en héptada en cada alfa-hélice. En cuarto lugar, las alfa-hélices deben estar estrechamente empaquetadas una contra otra por medio de sus cadenas laterales que interaccionan en forma de botones en ojales, tal como se ilustra para superhélices paralelas diméricas, triméricas y tetraméricas en Harbury *et al.* [Nature 1994, 371:80-83] y para superhélices antiparalelas triméricas en Lovejoy *et al.* [Science 1993, 259:1288-1293]. Walshaw *et al.* [J Struct Biol 2003, 144:349-361] describen normas más sofisticadas y un método

para distinguir las verdaderas superhélices de conjuntos de múltiples hélices.

Además de lo anterior, las moléculas de proteína de la presente invención existen como proteína aisladas y no requieren alfa-hélices asociadas adicionales (u otros fragmentos de proteína) para su plegamiento estable en disolución, como es el caso para determinadas clases de conjuntos de superhélice complejos enumerados en la "base de datos CC+ de superhélices" [[http://coiledcoils.chm.bris.ac.uk/ccplus/search/periodic table](http://coiledcoils.chm.bris.ac.uk/ccplus/search/periodic%20table)].

Además, la estructuras de superhélice de la presente invención no contienen irregularidades en sus secuencias de repetición en héptada (es decir, regiones entrecortadas o interrumpidas), lo que significa que tienen la separación de 3-4 convencional entre residuos de núcleo consecutivos (en las posiciones "a" y "d" de héptada convencionales) a lo largo de la secuencia.

Siempre que sea posible medir (es decir, si puede obtenerse una estructura 3-D), las moléculas de la presente invención tienen también un alto grado de simetría estructural, porque tienen capas separadas regularmente repetidas de residuos "a" de núcleo (capas a) y residuos "d" de núcleo (capas d) dentro de las dos alfa-hélices paralelas que existen dentro del plegamiento de superhélice antiparalela. Puesto que los residuos de núcleo forman los determinantes primarios del tipo de plegamiento, la simetría estructural también puede discernirse, e incluso imponerse, basándose en la secuencia de aminoácidos, es decir, mediante la selección apropiada de residuos de aminoácido de núcleo. Tal simetría estructural es importante para desarrollar moléculas de superhélice no naturales (diseñadas), porque hace que la tarea de diseño sea manejable (no pueden diseñarse *de novo* estructuras irregulares). Además, la creación de simetría estructural por medio de la introducción de simetría a nivel de los residuos de núcleo potencia considerablemente la probabilidad de plegamiento dando lugar a superhélices altamente estables, regulares.

Una posibilidad para garantizar la formación de capas a y d regulares es evitando la selección de residuos aromáticos voluminosos (triptófano, tirosina, fenilalanina) y residuos muy pequeños (glicina, alanina) en posiciones de núcleo. Otro modo de promover capas a y d regulares es seleccionando residuos de núcleo hidrófobos de tamaño moderado, tales como isoleucina, leucina, metionina y valina. Aún otro modo de obtener capas a y d regulares es seleccionando los mismos residuos de aminoácido en capas consecutivas del núcleo (por ejemplo, isoleucina en todas las posiciones de capa a). Aún otro modo de obtener capas de núcleo regulares es seleccionado los mismos residuos de aminoácido en posiciones de núcleo equivalentes en alfa-hélices adyacentes (por ejemplo, isoleucina en la posición a de la primera héptada tanto en la primera como en la tercera alfa-hélice, formando estas hélices las hélices paralelas de la estructura de superhélice). En general, cuanto mayor sea la simetría de la secuencia de aminoácidos en las posiciones de núcleo, mayor será la posibilidad de que las moléculas diseñadas se plieguen tal como se desea. Por tanto, las moléculas de la presente invención incluyen al menos algún grado, preferiblemente un grado razonable, lo más preferiblemente un alto grado de simetría de secuencia y, de ese modo, simetría estructural.

La existencia o falta de simetría forma un discriminador adecuado entre las moléculas de la presente invención y haces de 3 hélices conocidos que no forman las realizaciones de la invención. En la naturaleza, sólo se observan superhélices altamente simétricas como oligómeros y nunca como moléculas de una sola cadena. (Los motivos subyacentes para esta observación son complejos e intrigantes, pero son de poca importancia en el presente documento). A la inversa, los haces de 3 hélices de una sola cadena naturales no sólo son muy poco comunes (aparecen habitualmente como dominios antiparalelos pequeños en proteínas o complejos más grandes), sino que también carecen notablemente de simetría interna.

Uno de los ejemplos más cercanos de la técnica anterior sobre haces de 3 hélices antiparalelas se encuentra en la estructura de PDB del dominio GGA1 GAT humano [Zhu *et al.*, EMBO J 2004, 23:3909-3917; código de PDB: 1X79]. Los residuos 210-302 del dominio GGA1 GAT forman un motivo de haz de tres hélices antiparalelas que quizá podría confundirse con las estructuras de superhélice antiparalela de la presente invención. La primera alfa-hélice en este haz discurre en gran medida paralela a la tercera hélice, mientras que la segunda hélice se orienta antiparalela con respecto a estas. El empaquetamiento es relativamente apretado y se produce en un modo de botones en ojales. Sin embargo, las dos hélices paralelas carecen de simetría de empaquetamiento, tal como puede observarse a partir de la ausencia de capas a y d en la estructura, y la ausencia de residuos de núcleo de héptada estructuralmente similares en las secuencias de aminoácidos de las hélices 1 y 3: grandes (1 arginina, 1 tirosina) y pequeños (2 alaninas) que se imbrican con una mezcla residuos de núcleo alifáticos (valina, isoleucina, leucina). Además, los cristalógrafos no indican que el dominio GAT sea una superhélice (sino un haz de 3 hélices), mientras que clasifican el ligando rabaptina5 unido como una superhélice (dimérica). Otros ejemplos de haces de 3 hélices distintos de superhélice incluyen los dominios B, E y Z en la proteína A de estafilococos y la estructura terciaria de la parte anterior de villina.

El tipo y formato específicos de las moléculas que forman superhélice de la presente invención no se observan en la naturaleza, lo que es uno de los motivos por los que se denominan preferiblemente "no naturales".

La presente invención se refiere principalmente a, y una realización preferida de la presente invención incluye, una proteína de una sola cadena aislada que está representada por la fórmula HRS1-L1-HRS2-L2-HRS3, en la que

HRS1, L1, HRS2, L2 y HRS3 representan fragmentos de secuencia de aminoácidos que están interconectados covalentemente plegándose espontáneamente dicha proteína en disolución acuosa por medio de los fragmentos HRS1, HRS2 y HRS3 que forman una estructura de superhélice alfa-helicoidal, antiparalela, de triple hebra, y en la que

5 cada uno de HRS1, HRS2 y HRS3 es independientemente una secuencia de repetición en héptada que se caracteriza por un patrón de 7 residuos repetido n veces de tipos de aminoácidos, representado como (a-b-c-d-e-f-g)- n o (d-e-f-g-a-b-c)- n y una repetición en héptada parcial C-terminal que termina con un residuo de núcleo de héptada ubicado en una posición a o d, en la que los elementos de patrón "a" a "g" indican posiciones de héptada convencionales en las que se ubican dichos tipos de aminoácidos y n es un número igual a o mayor de 2, y

10 al menos el 50%, el 70%, el 90% o el 100% de las posiciones de héptada convencionales "a" y "d" están ocupadas por aminoácidos seleccionados del grupo que consiste en valina, isoleucina, leucina, metionina, fenilalanina, tirosina, triptófano, histidina, glutamina, treonina, serina, alanina o derivados de los mismos, en la que al menos el 50%, el 70%, el 90% o en la que el 100% de las posiciones de héptada convencionales "b", "c", "e", "f" y "g" están ocupadas por aminoácidos seleccionados del grupo que consiste en glicina, alanina, cisteína, serina, treonina, histidina, asparagina, ácido aspártico, glutamina, ácido glutámico, lisina, arginina o derivados de los mismos, permitiendo la distribución resultante de los tipos de aminoácidos la identificación de dichas secuencias de repetición en héptada, y

15 cada uno de L1 y L2 es independientemente un ligador que consiste en de 1 a 30 residuos de aminoácido, incluyendo este ligador cualquier residuo de aminoácido que no puede asignarse de manera inequívoca a una secuencia de repetición en héptada, y

20 L1 y L2 tienen una composición de aminoácidos que comprende al menos el 50% de aminoácidos seleccionados del grupo que consiste en glicina, alanina, serina, treonina, prolina o derivados de los mismos.

25 La propiedad mencionada anteriormente "aislado" se refiere esencialmente al requisito de que las proteínas de la presente invención formen estructuras estables sin necesidad de asociarse adicionalmente con ligandos (por ejemplo, otras proteínas, péptidos, ácidos nucleicos, hidratos de carbono, iones, etc.), o se embeban dentro de un contexto proteico mayor (es decir, dentro de un constructo de fusión o como un dominio), tal como se explicó también anteriormente.

30 La propiedad mencionada anteriormente "no natural" se refiere esencialmente al requisito de que las proteínas de la presente invención no se observen en la naturaleza, como proteínas naturales, o como dominios proteicos que se producen de manera natural. Para distinguirlas de los dominios o las proteínas naturales, el porcentaje de identidad de secuencia de aminoácidos equivale a preferiblemente menos del 90%, más preferiblemente menos del 80%, lo más preferiblemente menos del 70%. El término "no natural" también se refiere al hecho de que las proteínas se diseñan, o se conciben, preferiblemente según una base racional por humanos.

35 La propiedad mencionada anteriormente "una sola cadena" se refiere esencialmente al hecho de que las proteínas de la presente invención están hechas de una sola cadena de aminoácidos (cadena polipeptídica), y no de ensamblajes oligoméricos (diméricos, triméricos, etc.). Esto implica que pueden aislarse como monómeros en disolución. Sin embargo, esto último no excluye la posibilidad de que puedan interactuar con (asociarse con, formas complejos con, unirse a) otras moléculas, entidades biológicas o materiales no biológicos *in vitro* o *in vivo*.

40 La propiedad mencionada anteriormente "proteína" significa esencialmente un polipéptido compuesto por aminoácidos (residuos de aminoácido, opcionalmente aminoácidos no naturales o derivatizados) dispuestos en una cadena lineal y plegados en disolución (disolución acuosa, medio rico en agua) dando lugar a una forma globular.

45 El término mencionado antes "espontáneamente" significa esencialmente en un tiempo razonable (no extremo), en condiciones razonables, en sí mismo.

50 El término mencionado anteriormente "plegamiento" significa esencialmente la formación de un pliegue globular (compacto, "similar a un globo"), caracterizándose esta formación y estando dirigida por interacciones intracatenarias, interatómicas.

55 Los términos mencionados anteriormente "de triple hebra", "antiparalelo", "estructura de superhélice", "secuencia de repetición en héptada", "patrón", "posiciones de héptada convencionales", "predominantemente ocupado por", "tipos de aminoácidos hidrófobos", "tipos de aminoácidos hidrófilos", "ligador" y "asignado de manera inequívoca a una secuencia de repetición en héptada" tienen el significado explicado en otra parte en este documento. Se eligen para adaptarse de manera máxima a la terminología común en el campo.

60 También está relacionado con esta invención un método para la producción de dicha proteína. Un método de este tipo supone por ejemplo la expresión de dicha proteína en un huésped bacteriano, tal como se describe en el ejemplo 5. Alternativamente, la expresión de dicha proteína puede llevarse a cabo en sistemas eucariotas tales como células de levadura o insecto. Alternativamente, el pequeño tamaño de dicha proteína permite su producción

por medio de síntesis química, usando etapas de procedimiento bien conocidas en la técnica.

5 La presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que al menos el 50%, preferiblemente al menos 70%, al menos 90% o en la que el 100% (todas) las posiciones de héptada convencionales "a" y "d" están ocupadas por aminoácidos seleccionados del grupo que consiste en valina, isoleucina, leucina, metionina, fenilalanina, tirosina, triptófano, histidina, glutamina, treonina, serina, alanina o derivados de los mismos. El porcentaje preferido de dichos aminoácidos en dichas posiciones de héptada convencionales depende del nivel de riesgo que se está dispuesto a correr en el diseño de dicha proteína. Un porcentaje por debajo del 50% se considera que supone un riesgo demasiado alto como para que el plegamiento sea correcto.

15 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que al menos el 50%, el 70%, el 90% o en la que el 100% de las posiciones de héptada convencionales "a" y "d" están ocupadas por aminoácidos seleccionados del grupo que consiste en valina, isoleucina, leucina, metionina o derivados de los mismos. Puesto que estos últimos aminoácidos corresponden a residuos de núcleo de superhélice más convencionales (más frecuentemente observados), esta realización se prefiere con respecto a la anterior.

20 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que al menos el 50%, el 70%, el 90% o en la que el 100% de las posiciones de héptada convencionales "a" y "d" están ocupadas por isoleucinas. Puesto que el descubrimiento inicial de dicha proteína de superhélice antiparalela de una sola cadena se realizó con constructos que tienen residuos de isoleucina en posiciones de héptada convencionales "a" y "d", esta realización se prefiere con respecto a la anterior.

25 La presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que al menos el 50%, el 70%, el 90% o en la que el 100% de las posiciones de héptada convencionales "b", "c", "e", "f" y "g" están ocupadas por aminoácidos seleccionados del grupo que consiste en glicina, alanina, cisteína, serina, treonina, histidina, asparagina, ácido aspártico, glutamina, ácido glutámico, lisina, arginina o derivados de los mismos. El porcentaje preferido de dichos aminoácidos en dichas posiciones de héptada convencionales depende del nivel de riesgo que se está dispuesto a correr en el diseño de dicha proteína. Un porcentaje por debajo del 50% se considera que supone un riesgo demasiado alto como para que el plegamiento sea correcto y para la solubilidad de la proteína.

35 La presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que L1 y L2 tienen una composición de aminoácidos que comprende al menos el 50%, el 70%, el 90% o que comprende el 100% de aminoácidos seleccionados del grupo que consiste en glicina, alanina, cisteína, prolina, serina, treonina, histidina, asparagina, ácido aspártico, glutamina, ácido glutámico, lisina, arginina o derivados de los mismos. El porcentaje preferido de dichos aminoácidos dentro de los ligadores depende del nivel de riesgo que se está dispuesto a correr en el diseño de dicha proteína. Un porcentaje por debajo del 50% se considera que supone un riesgo demasiado alto como para que el plegamiento sea correcto, para la solubilidad de la proteína y para su posible función (por ejemplo, unión específica a una diana dada).

45 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que L1 y L2 tienen una composición de aminoácidos que comprende al menos el 50%, el 70%, el 90% o que comprende el 100% de aminoácidos seleccionados del grupo que consiste en glicina, alanina, serina, treonina, prolina o derivados de los mismos. Puesto que estos últimos aminoácidos corresponden a residuos de ligador más convencionales (seleccionados más habitualmente), esta realización se prefiere con respecto a la anterior.

50 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que L1 y L2 tienen una composición de aminoácidos que comprende al menos el 50%, el 70%, el 90% o que comprende el 100% de aminoácidos glicina y/o serina. Puesto que estos últimos aminoácidos corresponden a los residuos de ligador más convencionales (los más frecuentemente seleccionados), esta realización se prefiere con respecto a la anterior.

55 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que el número de residuos de aminoácido de cada uno de L1 y L2 equivale a menos de la mitad del número de residuos de aminoácido de la secuencia de repetición en héptada que precede al respectivo L1 o L2. Respetar esta norma disminuye considerablemente el riesgo de plegamiento no deliberado (por ejemplo, como una superhélice paralela), tal como se explicó anteriormente.

60 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que residuos de aminoácido cerca de los extremos terminales de L1 y/o L2 estabilizan los extremos alfa-helicoidales de la estructura de superhélice. Las posibilidades de seleccionar tales aminoácidos están bien documentadas en la bibliografía y se conocen generalmente como "aminoácidos de ocupación de extremos de hélice o motivos de ocupación de extremos de hélice".

65

- 5 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que residuos de aminoácido cerca de los extremos terminales de L1 y/o L2 promueven la formación de un giro local en la estructura. Las posibilidades de seleccionar tales aminoácidos incluyen, por ejemplo, la selección de aminoácidos que rompen la hélice tales como glicina y prolina, o aminoácidos que inician la hélice tales como serina o ácido aspártico. Determinados motivos de ocupación de extremos de hélice también se aplican para el mismo propósito. Alternativamente, pueden aplicarse motivos de hélice-bucle-hélice tal como se documenta en la bibliografía o se observa en el banco de datos de proteínas (PDB).
- 10 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que posiciones de héptada convencionales "e" y "g" están ocupadas por glutaminas. El modelado por ordenador de moléculas de superhélice antiparalela de la presente invención sugirió que pares de glutamina en dichas posiciones pueden formar interacciones casi ideales (es decir, enlaces de hidrógeno favorables energéticamente) entre hélices antiparalelas, aumentando de ese modo la estabilidad global del plegamiento.
- 15 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena en la que posiciones de héptada convencionales "b", "c" y "f" son aminoácidos polares que promueven la solubilidad. Puesto que estas posiciones son las más expuestas al disolvente, la selección exclusiva de aminoácidos polares, y preferiblemente cargados, en estas posiciones puede potenciar considerablemente la solubilidad de dicha proteína.
- 20 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena, que se pliega en disolución acuosa que tiene un pH de entre 1 y 13, o entre 2 y 12, o entre 3 y 11, o entre 4 y 10, o entre 5 y 9. El intervalo de pH en el que una proteína permanece plegada es un determinante importante de su aplicabilidad. Por ejemplo, la insensibilidad (tolerancia) a condiciones de pH extremas pueden hacerla adecuada para aplicaciones terapéuticas en las que la proteína necesita pasar a través de, o realizar su función en, el tracto gastrointestinal. Además, proteínas insensibles al pH pueden ser resistentes a las condiciones ácidas de la ruta lisosómica tras la endocitosis. Por tanto, las proteínas de la presente invención son preferiblemente estables en el intervalo de pH de 5-9, más preferiblemente 4-10, 3-11, 2-12 y lo más preferiblemente 1-13.
- 25 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena, que se pliega en disolución acuosa que tiene una temperatura de entre 0°C y 100°C, o entre 0°C y 80°C, o entre 0°C y 60°C. La estabilidad térmica es un determinante importante de la estabilidad global (incluyendo estabilidad proteolítica y estabilidad a largo plazo o "vida útil") y por tanto de la conservación de la función. Las proteínas de la presente invención son preferiblemente estables a intervalos temperatura de 0-60°C, más preferiblemente 0-80°C y lo más preferiblemente 0-100°C.
- 30 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena, que se pliega en disolución acuosa que tiene una fuerza iónica de entre 0 y 1,0 molar. Las condiciones fisiológicas requieren plegamiento estable y preservación de la función a fuerzas iónicas (en gran medida correspondientes a concentraciones salinas) de aproximadamente 150 milimolar. Las proteínas de la presente invención son preferiblemente estables (y funcionalmente activas) a intervalos más amplios de fuerza iónica, lo más preferiblemente en el intervalo de 0-1 molar.
- 35 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena, que se usa como almacén. Las moléculas de proteína de la presente invención son sumamente útiles como almacenes, tal como se explicó anteriormente.
- 40 Otra realización preferida de la presente invención se refiere a una proteína de superhélice antiparalela de una sola cadena, tal como se muestra en la figura 15.
- 45 Las proteínas de la presente invención son susceptibles de un gran número de modificaciones, usando el conocimiento de la técnica, incluyendo (múltiples) sustituciones de aminoácidos, introducción de aminoácidos no naturales, unión de restos químicos particulares, extensiones peptídicas, marcaje, potenciación de la aidez a través de autoconcatenación, concatenación en proteínas de fusión, etc., sin comprometer (cambiar, destruir) el plegamiento de superhélice de la proteína. Varias de tales modificaciones pueden formularse en el presente documento para ilustrar el potencial intrínseco de la proteína para someterse a etapas de modificación por ingeniería avanzadas. Concretamente, los presentes inventores contemplan los siguientes productos modificados por ingeniería, que incluyen todos la proteína de la presente invención con todas sus características especificadas:
- 50
- 55
- 60 • cualquier proteína de la presente invención puede modificarse en la secuencia de aminoácidos, creando de ese modo uno o más derivados de la misma;
 - cualquier proteína o derivado puede modificarse, por ejemplo, para mejorar su solubilidad;
 - 65 • cualquier proteína o derivado puede modificarse, por ejemplo, para mejorar su cinética de plegamiento;

- cualquier proteína o derivado puede modificarse, por ejemplo, para mejorar la corrección de su estado plegado;
- cualquier proteína o derivado puede modificarse, por ejemplo, para mejorar su afinidad de unión a un compuesto diana;
- cualquier proteína o derivado puede modificarse, por ejemplo, para mejorar su especificidad de unión por un compuesto diana;
- cualquier proteína o derivado puede modificarse, por ejemplo, para mejorar su solubilidad;
- cualquier proteína o derivado puede unirse covalentemente a cualquier otra proteína o molécula proteínica, o bien por medio de sus extremos N- y/o C-terminales o bien por medio de una o más de sus cadenas laterales;
- cualquier proteína o derivado puede unirse covalentemente a otras copias de la misma proteína o derivado, por ejemplo, para aumentar la avidéz;
- cualquier proteína o derivado puede unirse covalentemente a cualquier proteína o derivado con diferentes propiedades de unión, por ejemplo, para proporcionar bi o multiespecificidad;
- cualquier proteína o derivado puede unirse covalentemente a cualquier proteína o dominio proteico o péptido natural o no natural existente que no esté relacionado con la presente invención, incluyendo, sin limitación, dominios Fc, receptor de Fc, albúmina sérica, proteínas fluorescentes, moléculas de proteína de otro tipo, etc.;
- cualquier proteína o derivado puede unirse covalentemente a una o más etiquetas de detección;
- cualquier proteína o derivado puede unirse covalentemente a una o más etiquetas de purificación;
- cualquier proteína o derivado puede unirse covalentemente a compuestos orgánicos por medio de una reacción química con uno o más restos de cadena lateral de la proteína;
- cualquier proteína o derivado puede glicosilarse;
- cualquier proteína o derivado puede pegilarse.

En vista del hecho de que la proteína de la presente invención, y derivados de la misma, forman estructuras estables y compactas, pueden construirse o manipularse, en principio, mediante todas las técnicas aplicables a proteínas.

Las moléculas de proteína de la presente invención pueden prepararse de manera sintética según técnicas bien conocidas en la técnica o pueden producirse por medio de ingeniería genética usando técnicas que también se conocen bien en la técnica. Cuando se preparan con técnicas de ingeniería genética, las moléculas de proteína de la invención se codifican por polinucleótidos (también denominados en el presente documento ácidos nucleicos), preferiblemente ADN o ARN. Las moléculas de proteína de la invención pueden codificarse por cualquier ácido nucleico según la degeneración del código genético del organismo huésped en el que se prepara la molécula de proteína.

Los polinucleótidos (también denominados en el presente documento ácidos nucleicos) de la presente invención pueden incorporarse en un vector recombinante, por ejemplo un vector de clonación o expresión. El término "vector" incluye vectores de expresión, vectores de transformación y vectores lanzadera. El término "vector de expresión" significa un constructo capaz de expresión *in vivo* o *in vitro*. El término "vector de transformación" significa un constructo capaz de transferirse desde una entidad a otra entidad, que puede ser de la misma especie o puede ser de una especie diferente. Si el constructo es capaz de transferirse desde una especie hasta otra, tal como desde un vector viral tal como VLMM o VIF a una línea celular o célula primaria humana o de mamífero, entonces el vector de transformación se denomina algunas veces "vector lanzadera". Puede usarse una gran variedad de sistemas de expresión en diferentes huéspedes. Por ejemplo, sistemas episomales, cromosómicos y derivados de virus (por ejemplo vectores derivados de plásmidos bacterianos, bacteriófago, papovavirus tal como SV40, virus vaccinia, adenovirus y retrovirus). La secuencia de ADN puede insertarse en el vector mediante una variedad de técnicas. En general, la secuencia de ADN se inserta en un sitio de endonucleasa de restricción apropiado mediante procedimientos conocidos en la técnica y que se considera que están dentro del alcance de los expertos en la técnica. La secuencia de ADN en el vector de expresión está unida operativamente a secuencias de control apropiadas que dirigen la síntesis de ARNm (es decir, el promotor). Los vectores de la presente invención pueden transformarse en una célula huésped adecuada tal como se describe a continuación para proporcionar la expresión de una molécula de proteína de la presente invención. Por tanto, en un aspecto adicional, la invención proporciona un procedimiento para preparar moléculas de proteína según la presente invención que comprende cultivar una célula huésped transformada o transfectada con un vector de expresión tal como se describió anteriormente en condiciones para proporcionar la expresión mediante el vector de una secuencia codificante que codifica para las moléculas de proteína, y recuperar las moléculas de proteína expresadas. Los vectores pueden ser, por ejemplo,

vectores de plásmido, virus o bacteriófago (fago) dotados de un origen de replicación, opcionalmente un promotor para la expresión del polinucleótido y opcionalmente un regulador del promotor. Los vectores de la presente invención pueden contener uno o más genes marcadores seleccionables. Los sistemas de selección más adecuados para microorganismos industriales son los formados por el grupo de marcadores de selección que no requieren una mutación en el organismo huésped. Los ejemplos de marcadores de selección fúngicos son los genes para acetamidasa (*amdS*), ATP sintetasa, subunidad 9 (*oliC*), orotidina-5'-fosfato Descarboxilasa (*pvrA*), resistencia a fleomicina y benomilo (*benA*). Ejemplos de marcadores de selección no fúngicos son el gen de resistencia G418 bacteriano (también puede usarse en células de mamífero, levadura, pero no en hongos filamentosos), el gen de resistencia a ampicilina (*E. coli*), el gen de resistencia a neomicina (células de mamífero) y el gen *uidA* de *E. coli*, que codifica para beta-glucuronidasa (*GUS*). Pueden usarse vectores *in vitro*, por ejemplo para la producción de ARN o pueden usarse para transfectar o transformar una célula huésped. Por tanto, pueden incorporarse polinucleótidos o ácidos nucleicos de la presente invención en un vector recombinante (normalmente un vector que puede replicarse), por ejemplo un vector de clonación o expresión. El vector puede usarse para replicar el ácido nucleico en una célula huésped compatible. Por tanto, en una realización adicional, la invención proporciona un método de preparación de polinucleótidos de la presente invención introduciendo un polinucleótido de la presente invención en un vector que puede replicarse, introduciendo el vector en una célula huésped compatible y haciendo crecer la célula huésped en condiciones que provocan la replicación del vector. El vector puede recuperarse de la célula huésped. Se describen células huésped adecuadas a continuación en relación con vectores de expresión. El término "célula huésped", en relación con la presente invención, incluye cualquier célula que pueda comprender la secuencia de nucleótidos que codifica para la proteína recombinante según la presente invención y/o productos obtenidos a partir de la misma, en la que un promotor puede permitir la expresión de la secuencia de nucleótidos según la presente invención cuando está presente en la célula huésped. Por tanto, una realización adicional de la presente invención proporciona células huésped transformadas o transfectadas con un polinucleótido de la presente invención. Preferiblemente, dicho polinucleótido se porta en un vector para la replicación y expresión de dicho polinucleótido. Las células se elegirán para que sean compatibles con dicho vector y pueden ser, por ejemplo, procariontes (por ejemplo, células bacterianas) o eucariotes (es decir, células de mamífero, fúngicas, de insecto y de levadura). La introducción de polinucleótidos en células huésped puede efectuarse mediante métodos descritos en Sambrook, *et al.*, eds. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Nueva York, NY, EE.UU. Estos métodos incluyen, pero no se limitan a, transfección con fosfato de calcio, transfección mediada por DEAE-dextrano, transfección mediada por lípidos catiónicos, electroporación, transvección, microinyección, transducción, carga por raspado e introducción balística. Los ejemplos de huéspedes representativos incluyen células bacterianas (por ejemplo, *E. coli*, *Streptomyces*); células fúngicas tales como células de levadura y *Aspergillus*; células de insecto tales como células de *Drosophila* S2 y *Spodoptera* SF9; células animales tales como células CHO, COS, HEK, HeLa y 3T3. La selección del huésped apropiado se considera que está dentro del alcance de los expertos en la técnica. Dependiendo de la naturaleza del polinucleótido que codifica para la molécula de proteína de la presente invención, y/o el deseo de procesamiento adicional de la proteína expresada, pueden preferirse huéspedes eucariotes tales como levaduras u otros hongos. En general, se prefieren células de levadura con respecto a células fúngicas debido a que son más fáciles de manipular. Los ejemplos de huéspedes de expresión adecuados dentro del alcance de la presente invención son hongos tales como especies de *Aspergillus* y especies de *Trichoderma*; bacterias tales como especies de *Escherichia*, especies de *Streptomyces* y especies de *Pseudomonas*; y levaduras tales como especies de *Kluyveromyces* y especies de *Saccharomyces*. A modo de ejemplo, pueden seleccionarse huéspedes de expresión típicos de *Aspergillus niger*, *Aspergillus niger* var. *tubigenis*, *Aspergillus niger* var. *awamori*, *Aspergillus aculeatis*, *Aspergillus nidulans*, *Aspergillus oryzae*, *Trichoderma reesei*, *Kluyveromyces lactis*, *Schizosaccharomyces pombe*, *Pichia pastoris* y *Saccharomyces cerevisiae*. El uso de células huésped adecuadas, tales como células huésped de mamífero, de levadura, de insecto y fúngicas, puede proporcionar modificaciones postraduccionales (por ejemplo miristoilación, glicosilación, truncamiento y fosforilación de tirosina, serina o treonina) según pueda necesitarse para conferir una actividad biológica óptima a productos de expresión recombinantes de la presente invención. Tal como se indica, la célula huésped puede ser una célula procarionte o eucariota. Un ejemplo de un huésped procarionte adecuado es *E. coli*. Las enseñanzas sobre la transformación de huéspedes procariontes están bien documentadas en la técnica, por ejemplo véase Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, 2ª edición, 1989, Cold Spring Harbor Laboratory Press, Nueva York, NY, EE.UU.) y Ausubel *et al.* (*Current Protocols in Molecular Biology* (1995), John Wiley & Sons, Inc.). En una realización preferida, el huésped transformado es una célula de mamífero o, por ejemplo, una célula de insecto, en la que la introducción de polinucleótidos en dichas células huésped puede efectuarse mediante métodos descritos en, por ejemplo, Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, 2ª edición, 1989, Cold Spring Harbor Laboratory Press, Nueva York, NY, EE.UU.). Estos métodos incluyen, pero no se limitan a, transfección con fosfato de calcio, transfección mediada por DEAE-dextrano, transfección mediada por lípidos catiónicos, electroporación, transvección, microinyección, transducción, carga por raspado e introducción balística. En otra realización, el organismo transgénico puede ser una levadura. En este sentido, también se ha usado ampliamente levadura como vehículo para expresión génica heteróloga. La especie *Saccharomyces cerevisiae* tiene una larga historia de uso industrial, incluyendo su uso para expresión génica heteróloga. La expresión de genes heterólogos en *Saccharomyces cerevisiae* se ha revisado por Goodey *et al.* (1987, *Yeast Biotechnology*, D. R. Berry *et al.*, eds, págs. 401-429, Allen y Unwin, Londres) y por King *et al.* (1989, *Molecular and Cell Biology of Yeasts*, E. F. Walton y G. T. Yarronton, eds, págs. 107-133, Blackie, Glasgow). Según la presente invención, la producción de la molécula de proteína de la presente invención puede efectuarse mediante el cultivo de huéspedes de expresión eucariotes o procariontes, que se han transformado con uno o más polinucleótidos de la presente invención, en un medio de

5 fermentación de nutrientes convencional. La selección del medio apropiado puede basarse en la elección de huéspedes de expresión y/o basarse en los requisitos regulatorios del constructo de expresión. Tales medios los conocen bien los expertos en la técnica. El medio puede contener, si se desea, componentes adicionales que favorecen a los huéspedes de expresión transformados con respecto a microorganismos potencialmente contaminantes.

Ejemplos

Ejemplo 1. Secuencia de aminoácidos de un péptido sintético con residuos de núcleo y no de núcleo.

10 Este ejemplo proporciona la secuencia de aminoácidos de un péptido específico que está relacionado con la presente invención. La secuencia de aminoácidos, AIAAIQKQIAAIQKQIAAIQKQIA, se presenta en la notación de una sola letra, en la que A se refiere a alanina, I a isoleucina, Q a glutamina y K a lisina. Los péptidos con esta secuencia de aminoácidos forman complejos de superhélice alfa-helicoidal, de triple hebra por medio de sus
15 residuos de aminoácido de isoleucina y leucina que forman un núcleo hidrófobo (centro, interior) y los otros residuos que se orientan hacia el disolvente. El péptido artificial comprende tres repeticiones en héptada marcadas como "HR1", "HR2" y "HR3" en la figura 1.

20 La figura 1 es una representación esquemática de la secuencia de aminoácidos de un péptido artificial que comprende repeticiones en héptada (HRx), residuos de núcleo (recuadros negros), residuos no de núcleo (recuadros grises) y regiones flanqueantes (recuadros blancos). El péptido comprende además un residuo de núcleo de héptada C-terminal marcado como "t". El péptido comprende además fragmentos flanqueantes N- y C-terminales marcados como "N" y "C", respectivamente. Cada residuo de repetición en héptada se anota además con índices "a" a "g" y un número correspondiente al número de repetición en héptada. Los residuos de núcleo se ubican en las
25 posiciones en a y d. Los 6 residuos de núcleo de las tres repeticiones en héptada completas son isoleucinas. El residuo de isoleucina marcado como "a4" pertenece a la repetición en héptada parcial "t". Las repeticiones en héptada HR1, HR2 y HR3 y la repetición en héptada parcial "t" constituyen juntas una secuencia de repetición en héptada, que comienza con el residuo de núcleo a1 y termina con el residuo de núcleo a4.

Ejemplo 2. Principios de un complejo de superhélice alfa-helicoidal, de triple hebra.

30 Los residuos de núcleo de héptada se protegen del disolvente en complejos de superhélice alfa-helicoidal, de triple hebra, tal como se ilustra en la figura 2. Interacciones no covalentes entre residuos de núcleo en contacto (posiciones A y D en la figura 2) proporcionan la principal fuerza motriz termodinámica para que los péptidos adopten tal plegamiento.
35

40 La figura 2 es una representación en rueda helicoidal de estructuras de superhélice alfa-helicoidal, de triple hebra. El panel de la izquierda muestra una vista desde arriba de una superhélice paralela. El panel de la derecha muestra una vista desde arriba de una superhélice antiparalela. El panel del medio muestra la secuencia lineal de posiciones de repeticiones en héptada. Sólo se muestra una repetición en héptada por motivos de claridad. Se usan diferentes tonos para indicar posiciones topológicas específicas.

45 Los residuos de núcleo (posiciones A y D) están completamente escondidos en el complejo y no son accesibles para el disolvente. Los residuos no de núcleo (posiciones B, C, E, F y G) son accesibles al menos parcialmente para el disolvente (posiciones E, G menos que B, C, y posiciones B, C menos que F) y son susceptibles de sustituciones de aminoácido sin implicaciones (importantes) para la estabilidad del complejo.

Ejemplo 3. Estructura alfa-helicoidal y plegamiento/desplegamiento reversible.

50 Las superhélices alfa-helicoidales peptídicas no forman el objeto de la presente invención porque no se pliegan dando lugar a una proteína de una sola cadena. Sin embargo, las proteínas de una sola cadena de la presente invención sí comprenden una región de superhélice trimérica. Evidentemente, la conexión de los extremos N- y C-terminales mediante fragmentos de ligador puede influir (influirá) en la cinética de plegamiento, pero se espera que las propiedades físicas esenciales de los péptidos de superhélice "cortados" se conserven generalmente. Por tanto,
55 las superhélices peptídicas pueden servir como sistema de estudio. Para demostrar la formación cuantitativa de estructura secundaria alfa-helicoidal de un péptido artificial de referencia en disolución, los inventores han sintetizado el péptido con la secuencia de aminoácidos Ac-MSIEIQKQQAAIQKQIAAIQKQIYRMTP-NH2 y han registrado el espectro de dicroísmo circular (DC). La secuencia de aminoácidos se facilita en el código de una sola letra; Ac- y -NH2 significan que el péptido comenzaba por acetilo y terminaba en amida, respectivamente. Este péptido debe considerarse como un derivado del péptido de referencia compuesto por secuencia de repetición en
60 héptada triple (IAAIQKQ)3, con modificaciones en los extremos amino-(N-) y carboxilo-(C-) terminales para mejorar la naturaleza alfa-helicoidal de los extremos terminales (a menudo denominado ocupación de extremos). Más específicamente, los residuos flanqueantes Ac-MS- se unieron al extremo N-terminal, en combinación con la sustitución de dos residuos de alanina (AA) por dos residuos de ácido glutámico (EE) consecutivos en la primera
65 héptada de la secuencia de referencia. Además, los residuos flanqueantes -YRMTP-NH2 se unieron al extremo C-terminal, de manera que los aminoácidos isoleucina (I) y metionina (M) se ubican en las posiciones a y d de héptada

convencionales, lo que permite que esta secuencia flanqueante forme una héptada extra, aunque incompleta. La tirosina (Y) se introdujo en una posición b orientada hacia el disolvente para permitir la determinación espectrofotométrica de la concentración. Los residuos de arginina (R), treonina (T) y prolina (P-NH₂) se introdujeron para mejorar la ocupación de extremos helicoidales C-terminales. Además, el residuo de isoleucina (I) en la posición a de la segunda héptada se reemplazó por un residuo de glutamina (Q) para forzar que los péptidos que forman la superhélice se asocien del modo correcto (previsto), es decir, para determinar la formación de un complejo trimérico y para evitar posibles desviaciones del registro de héptadas [Eckert *et al.*, J Mol Biol 1998, 284:859-865].

Dicho péptido sintetizado se disolvió a una concentración de 292 microM en tampón fosfato (PBS) 20 mM, NaCl 150 mM, pH 7,2. Se midieron los espectros de DC entre 200 y 250 nm, a 5°C y 90°C (figura 3). El espectro a 5°C era indicativo de un alto contenido en estructura secundaria alfa-helicoidal, de acuerdo con lo esperado de que todas las regiones de héptada, pero no todos los residuos flanqueantes, se ensamblarían como superhélices alfa-helicoidales. El espectro a 90 grados Celsius mostró que la estructura alfa-helicoidal se perdía en gran medida, pero no completamente, a temperaturas elevadas.

Con el fin de analizar si la transición inducida por la temperatura entre estados helicoidales y no helicoidales era reversible, se realizó un barrido térmico hacia delante (ascendente) y hacia atrás (descendente) sobre la misma muestra, mediante el registro de la señal de DC a 222 nm en función de la temperatura a una velocidad de barrido de aproximadamente 1 grado Celsius por minuto (figura 4). Se observó que los barridos ascendentes y descendentes coincidían casi perfectamente, confirmando de ese modo el desplegamiento y replegamiento cuantitativos de los péptidos en la muestra.

Se analizó adicionalmente si la curva de desplegamiento térmico de la figura 4 se ajustaba a las ecuaciones termodinámicas que describen la reacción de plegamiento/desplegamiento en equilibrio entre tres moléculas de péptido libre (monomérico) y una entidad de complejo plegado (trimérico). Esta reacción se escribe generalmente como 3 péptido \rightleftharpoons péptido₃ en la que " \rightleftharpoons " se refiere a un equilibrio químico, "péptido" a un péptido monomérico en disolución y "péptido₃" a una entidad trimérica en el estado plegado (ensamblado, asociado). Esta curva de desplegamiento térmico se ajustó a las ecuaciones teóricas:

$$\theta(T) = \theta_M(T) + (\theta_T(T) - \theta_M(T)) \left(1 + \sqrt[3]{F \left(-\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{F}{27}} \right)} + \sqrt[3]{F \left(-\frac{1}{2} - \sqrt{\frac{1}{4} + \frac{F}{27}} \right)} \right)$$

en la que

$$F = \frac{\exp\left(-\frac{\Delta H_t}{RT} (1 - T/T_t) - \frac{\Delta C_p}{RT} (T - T_t - T \ln(T/T_t))\right)}{4}$$

y $T \equiv$ la temperatura, en grados Kelvin, de la muestra

$\theta(T) \equiv$ la señal de DC [θ]_{222 nm}, en grados cm² dmol⁻¹, en función de T

$\theta_M(T) \equiv$ la señal de DC para el péptido libre al 100% (monomérico) en función de T

$\theta_T(T) \equiv$ la señal de DC para el péptido asociado al 100% (trimérico) en función de T

$T_t \equiv$ la temperatura de transición, en la que el 50% de la concentración de péptido total está asociado

$\Delta H_t \equiv$ la diferencia de entalpía, en kJ por mol de péptido, entre estados mono y triméricos

$\Delta C_p \equiv$ la diferencia de capacidad calorífica, en J mol⁻¹ K⁻¹, entre estados mono y triméricos

$R \equiv$ la constante de gases ideales (universales) $\equiv 8,31$ J mol⁻¹ K⁻¹

En la figura 5 se muestran los resultados de esta operación de ajuste. Se encontró que la curva teórica coincidía casi perfectamente a lo largo de todo el intervalo de temperatura con la curva experimental, confirmando de ese modo la asociación trimérica de los péptidos.

La figura 5 representa el ajuste de una ecuación teórica para la asociación trimérica a los datos experimentales. Los datos experimentales se toman de la figura 4, curva marcada como "ASCENDENTE". Las ecuaciones teóricas

usadas se enumeraron anteriormente. Los parámetros ajustados (resultados de ajuste) se enumeran a la derecha de la figura 5. “T. de transición” corresponde a T_t , pero se expresa en grados Celsius. El parámetro “delta C_p ” se mantuvo constante a $3,0 \text{ kJ mol}^{-1} \text{ K}^{-1}$. Los parámetros “theta_M (T)” y “theta_T (T)” se trataron como funciones lineales de T, dando como resultado las líneas rectas discontinuas descritas por las respectivas desviaciones y pendientes indicadas a la derecha de la figura. “RMS Resid.” se refiere a la raíz cuadrática media de las diferencias entre puntos de datos experimentales y teóricos. La propia curva ajustada (teórica) se representa gráficamente en blanco en la figura y coincide a lo largo de todo el intervalo de temperatura con los puntos de datos experimentales mostrados en negro.

10 Ejemplo 4. Uso de todos los residuos de núcleo de isoleucina.

Para analizar si el residuo de glutamina en la posición a de la segunda héptada en el péptido de referencia del ejemplo 3 se requería para el plegamiento correcto (previsto) en una superhélice trimérica, se reemplazó este residuo por isoleucina, dando como resultado un péptido denominado “Q2a1” que tenía una secuencia con isoleucina en todas las posiciones de núcleo (excepto metionina dentro del fragmento flanqueante C-terminal). Para este fin, se sintetizó el péptido con la siguiente secuencia: Ac-MSIEIQKQIAAIQKQIAAIQKQIYRMTP-NH₂.

La figura 6 muestra la curva de desnaturalización térmica para una preparación de muestra del péptido Q2a1 en las mismas condiciones que en el ejemplo 3. La señal de DC global era algo inferior a la esperada, lo que podría deberse a una desviación instrumental, a un error en la determinación de la concentración, a una pureza inferior o a un contenido en alfa-hélice inferior al esperado. No obstante, el principal objetivo de este experimento era examinar el efecto del mutante glutamina a isoleucina sobre la estabilidad del complejo. Por tanto resultó interesante hallar que esta variante mostraba una resistencia extremadamente alta frente a la desnaturalización térmica, es decir, era extremadamente termoestable. La temperatura de transición estimada era de alrededor de 97 grados Celsius, aunque esta última era difícil de determinar debido a que no se completó la transición. Además, el barrido descendente mostró una recuperación completa de la señal de DC, lo que indica reversibilidad completa.

Para confirmar que el complejo ensamblado tenía el peso molecular (PM) correcto, tal como se esperaba para un trímero, se sometió el péptido Q2a1 a ultracentrifugación en equilibrio de sedimentación analítica a 25000 rpm a una concentración de aproximadamente 1 mg/ml. La figura 7 muestra la curva de densidad óptica (DO) linealizada en comparación con las curvas teóricas para complejos monoméricos, diméricos y triméricos. Se encontró que los puntos de datos experimentales coincidían muy bien con la curva de modelo trimérico. A partir de la pendiente de la línea de regresión lineal, se dedujo el peso molecular aparente de 10500 Da, que concordaba bien con el valor teórico de 10242 Da (3 veces el PM de 3414 Da para un monómero).

Para confirmar adicionalmente la formación de complejos triméricos, se analizó el mismo péptido Q2a1 mediante dispersión de luz estática. Se pusieron 200 microlitros de péptido a 1 mg/ml en PBS en una columna de filtración en gel Superdex 75 10/300 GL conectada a detectores de ultravioleta (UV), de índice de refracción (RI) y de dispersión de luz estática (SLS). La figura 8 muestra los resultados. Las señales (curvas) de los tres detectores diferentes se marcan en consecuencia. Se observó un pico de dispersión de luz bien conformado que coincidía con un pico de UV y RI. El peso molecular aparente derivado del pico de UV era de 12530 ± 1510 Da, de nuevo concordando bien con el valor esperado.

Se concluyó que el uso de residuos de núcleo todos de isoleucina no tenía ningún efecto adverso sobre el ensamblaje de los péptidos en trímeros, tal como podría esperarse basándose en las consideraciones teóricas sobre las posibles desviaciones de registro de héptadas (no previstas). En cambio, todas las pruebas indicaron el plegamiento apropiado y exclusivo dando lugar a trímeros con el peso molecular correcto (esperado). Además, este péptido de núcleo todo de isoleucina tenía una estabilidad térmica muy alta, por lo que no se desplegaba cuantitativamente hasta 95 grados Celsius. Por tanto, este péptido puede considerarse como un péptido que forma superhélice trimérica preferido.

50 Ejemplo 5. Constructos de armazón de superhélice de una sola cadena.

Con el fin de examinar si podrían derivarse armazones de superhélice de una sola cadena de superhélices peptídicas por medio de la conexión de extremos terminales de secuencias de repetición en héptada (HRS) individuales usando fragmentos de ligador estructuralmente flexibles, se diseñaron tres constructos con diferentes longitudes de ligador, se produjeron y se sometieron a prueba. Concretamente, se construyeron las moléculas de armazón de superhélice de una sola cadena con las secuencias de aminoácidos enumeradas en la figura 9. Estos armazones se derivaron del armazón de superhélice trimérica peptídica del ejemplo 4 (Q2a1). Se sometieron a prueba ligadores ricos en Gly/Ser de 8 y 16 aminoácidos de longitud. Estos constructos se indican en el presente documento como “scQ2a1_L8” y “scQ2a1_L16”, respectivamente. En vista de la definición de secuencias de repetición en héptada (proporcionada anteriormente) que comienzan y terminan con un residuo de núcleo, se incluyen formalmente los residuos de ocupación de extremos N- y C-terminales metionina-serina (“MS”) y treonina (“T”), respectivamente, en los ligadores, y las secuencias “MGHHHHHHHHSSGHIEGRHMS” y “TP” se consideran como secuencias flanqueantes. La secuencia flanqueante N-terminal (secuencia líder) comprende una cola de 10-His (HHHHHHHHHH) seguida por un sitio de escisión de “factor Xa” (IEGRH).

Se produjeron los constructos según el siguiente método. Se recuperaron genes que codificaban para los constructos. Se optimizaron las secuencias de nucleótidos para que coincidieran con el uso de codones para la expresión en *E. coli*. Se proporcionaron los genes en el plásmido pCR 4 TOPO y se les unió un sitio de restricción NdeI en 3' y uno XhoI en 5' para la posterior subclonación en el vector pET16b (Novagen). Se transformó este último en la cepa BL21(DE3)/pLysE de *E. coli* y se realizaron pruebas de expresión a pequeña escala. Brevemente, se inocularon 25 ml de medio que contenía los antibióticos apropiados (medio LB + ampicilina 50 microg/ml + cloranfenicol 25 microg/ml) con un cultivo durante la noche (dilución 1/150x) y se hicieron crecer las células a 37°C hasta que DO600 alcanzó aproximadamente 0,65. Entonces se indujo la expresión de las proteínas diana mediante la adición de 0,4 mM de IPTG y se hicieron crecer adicionalmente las células a o bien 37°C o bien 30°C. Se tomaron alícuotas de cultivo tras 3,5 horas (t1, 37°C) y 5,5 horas (t2, 37°C y t2', 30°C) y se analizaron en geles de SDS-PAGE (acrilo al 10%, tinción de Coomassie), junto con una muestra antes de la inducción (t0). Para todos los constructos, tras la inducción, apareció una banda a aproximadamente el PM esperado.

Para aislar proteína de la fracción soluble, se indujeron aproximadamente 1,3 litros de cultivo durante 5,5 horas a 30°C. Se recogieron las células, se resuspendieron en un tampón Tris 50 mM, NaCl 150 mM, pH 7,8 y luego se rompieron haciéndolas pasar a través de un pulverizador celular. Se recuperó la fracción soluble mediante centrifugación y se cargó sobre una columna de 5 ml cargada con Ni²⁺ para el aislamiento basado en IMAC de la proteína diana. Se lavó la columna con 10 volúmenes de columna de tampón que contenía 20 mM de imidazol y se usó un gradiente de 20 a 600 mM de imidazol para la etapa de elución. Se reunieron las fracciones que contenían proteína y se concentraron desde ~15 hasta ~6 ml (Vivaspin MWCO 5 kDa, 2800 rpm). Se purificaron adicionalmente las proteínas sobre una columna de filtración en gel preparativa (Superdex 75 16/90; Tris 50 mM, NaCl 150 mM, pH 7,8 como tampón de transferencia; dos series; ~3 ml cargados/serie). Las proteínas eluyeron a alrededor de 130 ml; se reunieron las fracciones relevantes y se concentraron hasta un volumen final de ~10 ml (Vivaspin MWCO 5 kDa, 2800 rpm). Los niveles de expresión soluble calculados estaban en el intervalo de 10 - 15 mg por litro de cultivo bacteriano.

La figura 10 muestra el barrido térmico de DC para el constructo scQ2aI_L16 en PBS 20 mM, NaCl 150 mM, pH 7,2. El barrido térmico indica que no hay desplegamiento térmico hasta 90 grados Celsius. Esto muestra que dicho constructo es hipertermostable, con una temperatura de transición que excede de 100 grados Celsius.

Para poder observar una transición completa, se realizaron experimentos de desplegamiento térmico posteriores en presencia de clorhidrato de guanidinio 6 M (GuHCl). La figura 11 muestra los barridos de desnaturalización térmica de scQ2aI_L16 y scQ2aI_L8 en GuHCl 6 M registrados mediante DC a 222 nm. La concentración de proteína era de aproximadamente 30 μM en el mismo tampón PBS. Se ajustaron los barridos a un modelo de transición de dos estados y se convirtieron en fracción de proteína plegada. Se encontró que la temperatura de transición del constructo scQ2aI_L8 era de 7 grados Celsius más alta que la del constructo scQ2aI_L16. Este resultado no se esperaba porque sólo se suministra al constructo L16 ligadores que sean lo suficientemente largos como para formar un puente en la distancia entre los extremos terminales helicoidales en orientación paralela ("conexión hacia abajo"). Tal como se describió anteriormente, para una conexión hacia abajo, el número de residuos en el ligador debe ser al menos la mitad del número de residuos en las hélices en superhélice. De hecho, el ligador de Gly/Ser-ligador de 8 residuos comprende menos de los $28/2 = 14$ residuos que se requieren teóricamente, incluso si se toman los residuos de ocupación de extremos como parte del ligador (es decir, ignorando el hecho de que es necesario que permitan la reversión de la dirección de la cadena, lo que también requiere al menos uno o dos residuos). Por tanto, se concluyó que la termoestabilidad superior del constructo scQ2aI_L8 estaba en contradicción con una estructura de superhélice paralela.

Se consideró también que los ligadores demasiado cortos podrían inducir desplegamiento local de uno o más de los extremos terminales helicoidales, y de ese modo permitirían todavía un cierre hacia abajo en orientación paralela. Esta hipótesis se consideró improbable porque tal fenómeno produciría lógicamente un constructo menos estable en lugar de la estabilidad superior observada. No obstante, con el fin de excluir esta última posibilidad, se preparó una serie de constructos "cortos" que comprenden una héptada menos en cada alfa-hélice. Concretamente, las secuencias de repetición en héptada de los nuevos constructos consistía en la secuencia IEEIQKQIAAIQKQIYRM (en lugar de IEEIQKQIAAIQKQIAAIQKQIYRM), con segmentos flanqueantes por lo demás idénticos y ligadores de Gly/Ser de fórmula (GGSG)_nGG con n = 1, 2, 3, 4, produciendo los respectivos constructos denominados "short_L6", "short_L10", "short_L14" y "short_L18". Se razonó que, si se produjera desplegamiento local para los constructos con ligadores demasiado cortos (teóricamente, para los constructos L6 y L10), esto debería reducir definitivamente su estabilidad térmica. Por tanto, se sometieron a prueba estos constructos mediante barrido térmico de DC a concentraciones variables de GuHCl, y se determinaron sus temperaturas de transición. La figura 12 muestra los resultados. Se encontró que los cuatro constructos cortos eran menos estables que el scQ2aI_L16 de referencia en aproximadamente 40 grados Celsius a las mismas concentraciones de GuHCl, lo que se esperaba en vista de los tamaños de superhélice reducidos. Las estabildades relativas de los cuatro constructos cortos eran altamente similares en todas las condiciones sometidas a prueba. A la concentración de GuHCl más alta (4 M), el constructo con el ligador más corto (short_L6) era de nuevo un poco más estable que los otros. Por tanto, se concluyó que la hipótesis de desenrollamiento helicoidal local no se cumple y, por tanto, que lo más probablemente todos los constructos no son paralelos sino, en cambio, antiparalelos.

Ejemplo 6. Experimentos de RMN.

5 Para proporcionar adicionalmente evidencias del plegamiento antiparalelo de las secuencias de superhélice de referencia de los ejemplos anteriores, se registraron los espectros de RMN de ^{15}N ^1H HSQC para los constructos scQ2a1_L16 y scQ2a1_L8. La figura 13 muestra los espectros, marcados como "L16" y "L8", respectivamente. Las amidas de cadenas laterales y estructura principal se agrupan en el cuadrante superior derecho e inferior izquierdo, respectivamente, y las amidas de estructura principal de ligador más flexible se agrupan en el cuadrante superior izquierdo. Se observa que los dos espectros son altamente similares, lo que es indicativo de un tipo de plegamiento que es independiente de la longitud de ligador. Puesto que el ligador L8 es estructuralmente incompatible con el plegamiento paralelo, se concluyó a partir de estos resultados que ambos son lo más probablemente antiparalelos.

15 Para proporcionar evidencias adicionales, se preparó un derivado de scQ2a1_L16 en el que se introdujo un triptófano (W) cerca del extremo N-terminal de la segunda hélice y una cisteína (C) cerca del extremo C-terminal de la tercera hélice. La secuencia de aminoácidos completa era MGHHHHHHHHHSSGHIEGRHMS-IEEIQKQIAAIQKQIAAIQKQIYRM-TGGSGGGSGGGSGGGSG~~WS~~-IEEIQKQIAAIQKQIAAIQKQIYRM-TGGSGGGSGGGSGGGSGMS-IEEIQKQIAAIQKQIAAIQ~~C~~QIYRM-TP (mutaciones resaltadas). Si esta secuencia se pliega como una superhélice antiparalela de una sola cadena, entonces las dos posiciones mutadas deben estar próximas en el espacio. Esto último puede comprobarse por medio de la conjugación de la cisteína con un marcador de espín y la monitorización del efecto del marcador de espín sobre la resonancia de la cadena lateral de triptófano NH ϵ . Si la cisteína marcada y el triptófano están en proximidad estrecha (es decir, preferiblemente menos de aproximadamente 15 Å), entonces la señal de triptófano NH ϵ debe disminuir significativamente. El tratamiento con vitamina C reduce el radical libre NO y de ese modo restaura (o aumenta) la señal de NH ϵ .

25 La figura 14 muestra las resonancias de RMN de dicho NH ϵ de triptófano de dicho constructo mutado. El marcador de espín usado en el presente experimento era 3-(2-yodoacetamido)-proxilo [es decir, 3-(2-yodoacetamido)-2,2,5,5 tetrametil-1-pirrolidinóxilo, radical libre de Acros Organics, n.º de cat. 224980250]. Cuando se comparan las señales para las muestras no tratadas y tratadas con vitamina C (marcadas en consecuencia en la figura 14), se observa que la señal de la muestra no tratada, que lleva el marcador de espín de radical libre, está de hecho significativamente disminuida en comparación con la muestra de control con el marcador reducido. Esto demuestra que el triptófano y la cisteína están en proximidad estrecha, lo que, en vista de las dimensiones de la estructura de superhélice (aproximadamente 40 Å de longitud), sólo es posible en un plegamiento antiparalelo.

Ejemplo 7. Modelado molecular de superhélices de una sola cadena paralelas y antiparalelas

35 La figura 15 representa modelos moleculares 3-D de una superhélice de una sola cadena de triple hebra paralela (panel izquierdo) y antiparalela (panel derecho) con la secuencia de aminoácidos del constructo scQ2a1_L16 (sin cola N-terminal). Las alfa-hélices constituidas de HRS1, HRS2 y HRS3 se indican respectivamente como A, B y C. Los dos fragmentos de ligador se marcan como L1 y L2, respectivamente.

40 Se construyó el modelo paralelo mediante modelado de homología partiendo de la estructura de PDB 1GCM. Se construyó el modelo antiparalelo mediante la inversión de la orientación de la hélice B en el modelo paralelo, seguido por su desviación a lo largo de su eje helicoidal hasta que todas las cadenas laterales estaban libres de solapamiento atómico. Esto último se logró sin modificar las estructuras rotaméricas de las cadenas laterales del núcleo. Se modelaron fragmentos de ligador mediante una combinación de rotación interactiva alrededor de los ángulos diédricos de la cadena principal, simulaciones de dinámica molecular y minimizaciones de energía, al mismo tiempo que se restringían los segmentos alfa-helicoidales.

50 Los modelos se han generado para examinar si la orientación antiparalela es estructuralmente factible. Puesto que todas las cadenas laterales que forman el núcleo pudieron colocarse en su conformación rotamérica más relajada, sin dejar cavidades intermitentes, y dando como resultado un empaquetamiento creíble de cada capa de héptada, se concluyó que la orientación antiparalela es estructuralmente posible, al menos en los modelos mostrados.

TABLA 1

Clase	Plegamiento	Superfamilia	Proteína	Especie	Código de PDB
Superhélice	Paralelo	Dominio de superhélice triple de lectinas de tipo C	De unión a manosa	Humano	1HUP
			Tensioactiva	Rata	1BUU, 1AFA, 1AFB, 1AFD, 1BCH, 1BCJ, 1FIF, 1FIH, 1KMB, 1KWT, 1KWJ, 1KWV, 1KWW, 1KWX, 1KWY, 1KWZ, 1KXO, 7KX1, 1RTM, 2KMB, 3KMB, 4KMB
				Humano	1PWB, 1PW9, 1B08, 1M7L, 2GGU, 2GGX, 2ORJ, 2ORK, 2OS9
			Tetranectina	Rata	1R13, 1R14
				Humano	1HTN
		Dominio de trimerización de TRAF	TRAF2	Humano	1D01, 1CA4, 1CA9, 1CZY, 1CZZ, 1D00, 1D0A, 1D0J, 1F3V, 1QSC
			TRAF3	Humano	1L0A, 1FLK, 1FLL, 1KZZ, 1RF3, 1ZMS, 2GKW
		Dominio de cremallera de leucina	GCN4	Levadura	1PIQ, 1GCM, 1ZIM, 1IJ3, 1IJ2, 1IJ1, 1IJ0, 1SWI, 1ZJ, 1CE0, 1EBO, 1ENV, 1FAV, 2B9B, 1CZQ, 1GZL, 2Q3I, 2Q5U, 2Q7C, 2R3C, 2R5B, 2R5D, 2OXJ
		Matriz de cartilago de pollo		Pollo	Antiparalelo; 1RB1
		Lipoproteína de la membrana externa			1AQ5
		Fibrina		<i>E. coli</i>	1EQ7, 1KFM, 1KFN, 1JCC, 1JCD
		Similar a MPN010	MPN010	Bacteriófago T4	1AA0, 2BSG, 1AVY, 1OX3, 2IBL
		Coronina 1		<i>Mycoplasma pneumoniae</i>	2BA2
		DMPK		Ratón	2AKF
				Humano	1WT6
	Segmentos de tallo de proteínas de fusión víricas	Hemaglutinina de influenza		Influenza A	1QU1, 1EO8, 1HA0, 1HGD, 1HGE, 1HGF, 1HGG, 1HGH, 1HGI, 1HGJ, 1KEN, 1QFU, 2HMG, 2VIU, 3HMG, 4HMG, 5HMG, 1MQL, 1MQM, 1MQN, 1HTM, 1TI8, 1RD8, 1RUZ, 1RV0, 1RVT, 1RUY, 2FK0, 2IBX, 1RU7, 1RVX, 1RVZ
				Influenza C	1FLC
		Ectodominio de virus	Retrovirus gp41		1DF4, 1AIK, 1DF5, 1DLB, 1ENV, 1FAV, 1I5X, 1K33, 1K34, 1S2T, 2CMR, 1F23, 1QR9, 1I5Y, 1QR8, 1CZO, 1GZL, 2Q3I, 2Q5U, 2Q7C, 2R3C, 2R5B, 2R5D
				VIH tipo 1	1QBZ, 1QCE, 2EZO, 2EZP, 2EQZ, 2EZR, 2EZS, 1JPX, 2SIV, 1JQD
				VIS	

						Visna		1JEK	
			gp21 de VLHT-1			VLHT tipo 1		1MG1	
			gp2 de Ebo			virus Ébola		2EBO, 1EBO	
			p15 de VLMmO			VLMmO		1MOF	
			sv5 de paramixovirus			cepa SV5 w3		1SVF	
			hPIV3 de paramixovirus			cepa hPIV3		1ZTM	
			Virus de las paperas			Virus de las paperas		2FYZ	
			Tallo de NDV			NDV		1G5G	
			Fusión de VRSH			VRSH		1G2C	
			HERV-FRD			Humano		1Y4M	
			Virus Nipah			Virus Nipah		1WP7	
			Virus Hendra			Virus Hendra		1WP8	
		Coronavirus S2	Espiga de E2			VHM		1WDG, 1WDF	
						SARS		2BEQ, 1WNC, 1WYY, 1ZV8, 2BEZ, 1ZVB, 2FXP	
						NL63		2IEQ	
						Sintético		2JGO	
Diseñada	Serina enrollada					Sintético		Antiparalelo: 1COS	
	Superhélice trimérica diseñada					Sintético		1COI	
			Vald			Sintético		Antiparalelo: 1G6U	
			Diseño amiloideogénico 1			Sintético		1S9Z	
			Diseño no nombrado 1			Sintético		1HQJ	
			Diseño no nombrado 2			Sintético		1KYC	
			Trímero de superhélice dextrógira			Sintético		1TGG	
Todas proteínas alfa	Superhélice dextrógira					Sintético			
	Proteína YhaI hipotética					<i>Bacillus subtilis</i>		1SED	
Membrana y superficie celular	Dominio de interacción con membrana VP4					Rotavirus de mono rhesus		1SLQ	
Proteínas pequeñas	Resistina					Ratón		1RGX, 1RFX	
			Similar a resistina			Ratón		1RH7	
		Dominio de unión a proteína G	Rabaptina-5			Humano		Antiparalelo: 1X79	

Lista de secuencias

- <110> Complix NV
- 5 <120> Proteínas de superhélice antiparalela de una sola cadena
- <130> 138956
- <160> 23
- 10 <170> PatentIn versión 3.4
- <210> 1
- <211> 24
- 15 <212> PRT
- <213> Artificial
- <220>
- <223> Péptido sintético
- 20 <400> 1
- Ala Ile Ala Ala Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile
- 1 5 10 15
- Ala Ala Ile Gln Lys Gln Ile Ala
- 20
- <210> 2
- 25 <211> 29
- <212> PRT
- <213> Artificial
- <220>
- 30 <223> Péptido sintético
- <220>
- <221> MISC_FEATURE
- <222> (1)..(1)
- 35 <223> Extremo N-terminal acetilado
- <220>
- <221> MISC_FEATURE
- <222> (29)..(29)
- 40 <223> Extremo C-terminal amidado
- <400> 2
- Met Ser Ile Glu Glu Ile Gln Lys Gln Gln Ala Ala Ile Gln Lys Gln
- 1 5 10 15
- Ile Ala Ala Ile Gln Lys Gln Ile Tyr Arg Met Thr Pro
- 20 25
- 45 <210> 3
- <211> 29
- <212> PRT
- <213> Artificial
- 50 <220>
- <223> Péptido sintético
- <220>
- <221> MISC_FEATURE

ES 2 547 121 T3

<222> (1)..(1)
 <223> Extremo N-terminal acetilado

<220>
 5 <221> MISC_FEATURE
 <222> (29)..(29)
 <223> Extremo C-terminal amidado

<400> 3
 Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln
 1 5 10 15

10 Ile Ala Ala Ile Gln Lys Gln Ile Tyr Arg Met Thr Pro
 20 25

<210> 4
 <211> 122
 <212> PRT
 15 <213> Artificial

<220>
 <223> Péptido sintético (scQ2al_L8)

20 <400> 4
 Met Gly His His His His His His His His His His His Ser Ser Gly His
 1 5 10 15

Ile Glu Gly Arg His Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala
 20 25 30

Ala Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Tyr Arg Met
 35 40 45

Thr Gly Gly Ser Gly Gly Gly Ser Gly Met Ser Ile Glu Glu Ile Gln
 50 55 60

Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln
 65 70 75 80

Ile Tyr Arg Met Thr Gly Gly Ser Gly Gly Gly Ser Gly Met Ser Ile
 85 90 95

Glu Glu Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Ala Ala
 100 105 110

Ile Gln Lys Gln Ile Tyr Arg Met Thr Pro
 115 120

<210> 5
 <211> 138
 <212> PRT
 25 <213> Artificial

ES 2 547 121 T3

<220>

<223> Péptido sintético (scQ2a_IL16)

5

<400> 5

Met Gly His His His His His His His His His His Ser Ser Gly His
1 5 10 15

Ile Glu Gly Arg His Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala
20 25 30

Ala Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Tyr Arg Met
35 40 45

Thr Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
50 55 60

Gly Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys
65 70 75 80

Gln Ile Ala Ala Ile Gln Lys Gln Ile Tyr Arg Met Thr Gly Gly Ser
85 90 95

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Met Ser Ile
100 105 110

Glu Glu Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Ala Ala
115 120 125

Ile Gln Lys Gln Ile Tyr Arg Met Thr Pro
130 135

<210> 6

<211> 97

10

<212> PRT

<213> Artificial

<220>

<223> Péptido sintético (short_L6)

15

<400> 6

Met Gly His His His His His His His His His His Ser Ser Gly His
1 5 10 15

Ile Glu Gly Arg His Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala
20 25 30

Ala Ile Gln Lys Gln Ile Tyr Arg Met Thr Gly Gly Ser Gly Gly Gly

ES 2 547 121 T3

35

40

45

Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln
50 55 60

Ile Tyr Arg Met Thr Gly Gly Ser Gly Gly Gly Met Ser Ile Glu Glu
65 70 75 80

Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Tyr Arg Met Thr
85 90 95

Pro

5 <210> 7
<211> 105
<212> PRT
<213> Artificial

10 <220>
<223> Péptido sintético (short_L10)

<400> 7
Met Gly His His His His His His His His His His Ser Ser Gly His
1 5 10 15

Ile Glu Gly Arg His Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala
20 25 30

Ala Ile Gln Lys Gln Ile Tyr Arg Met Thr Gly Gly Ser Gly Gly Gly
35 40 45

Ser Gly Gly Gly Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala Ala
50 55 60

Ile Gln Lys Gln Ile Tyr Arg Met Thr Gly Gly Ser Gly Gly Gly Ser
65 70 75 80

Gly Gly Gly Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala Ala Ile
85 90 95

Gln Lys Gln Ile Tyr Arg Met Thr Pro
100 105

15 <210> 8
<211> 113
<212> PRT
<213> Artificial

<220>

ES 2 547 121 T3

<223> Péptido sintético (short_L14)

<400> 8

Met Gly His His His His His His His His His His His Ser Ser Gly His
1 5 10 15

Ile Glu Gly Arg His Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala
20 25 30

Ala Ile Gln Lys Gln Ile Tyr Arg Met Thr Gly Gly Ser Gly Gly Gly
35 40 45

Ser Gly Gly Gly Ser Gly Gly Gly Met Ser Ile Glu Glu Ile Gln Lys
50 55 60

Gln Ile Ala Ala Ile Gln Lys Gln Ile Tyr Arg Met Thr Gly Gly Ser
65 70 75 80

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Met Ser Ile Glu Glu
85 90 95

Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Tyr Arg Met Thr
100 105 110

Pro

5

<210> 9

<211> 121

<212> PRT

<213> Artificial

10

<220>

<223> Péptido sintético (short_L18)

<400> 9

ES 2 547 121 T3

Met Gly His His His His His His His His His His Ser Ser Gly His
 1 5 10 15

Ile Glu Gly Arg His Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala
 20 25 30

Ala Ile Gln Lys Gln Ile Tyr Arg Met Thr Gly Gly Ser Gly Gly Gly
 35 40 45

Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Met Ser Ile Glu
 50 55 60

Glu Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Tyr Arg Met
 65 70 75 80

Thr Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
 85 90 95

Gly Gly Gly Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala Ala Ile
 100 105 110

Gln Lys Gln Ile Tyr Arg Met Thr Pro
 115 120

- <210> 10
- 5 <211> 138
- <212> PRT
- <213> Artificial

- <220>
- 10 <223> Péptido sintético (derivado de scQ2a1_L16)

<400> 10

ES 2 547 121 T3

Met Gly His His His His His His His His His His Ser Ser Gly His
1 5 10 15

Ile Glu Gly Arg His Met Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala
20 25 30

Ala Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Tyr Arg Met
35 40 45

Thr Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
50 55 60

Gly Trp Ser Ile Glu Glu Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys
65 70 75 80

Gln Ile Ala Ala Ile Gln Lys Gln Ile Tyr Arg Met Thr Gly Gly Ser
85 90 95

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Met Ser Ile
100 105 110

Glu Glu Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Ala Ala
115 120 125

Ile Gln Cys Gln Ile Tyr Arg Met Thr Pro
130 135

<210> 11

<211> 7

5 <212> PRT

<213> Artificial

<220>

10 <223> Secuencia de repetición en héptada

<400> 11

Ile Ala Ala Ile Gln Lys Gln

1 5

<210> 12

15 <211> 6

<212> PRT

<213> Artificial

<220>

20 <223> Secuencia flanqueante

<220>

<221> MISC_FEATURE

<222> (6)..(6)

25 <223> Extremo C-terminal amidado

ES 2 547 121 T3

<400> 12
 Ile Tyr Arg Met Thr Pro
 1 5

5 <210> 13
 <211> 23
 <212> PRT
 <213> Artificial

10 <220>
 <223> Secuencia flanqueante

<400> 13
 Met Gly His His His His His His His His His His Ser Ser Gly His
 1 5 10 15

Ile Glu Gly Arg His Met Ser
 20

15 <210> 14
 <211> 10
 <212> PRT
 <213> Artificial

20 <220>
 <223> Cola de 10-His

<400> 14
 His His His His His His His His His His
 1 5 10

25 <210> 15
 <211> 5
 <212> PRT
 <213> Artificial

30 <220>
 <223> Sitio de escisión de factor Xa

<400> 15
 Ile Glu Gly Arg His
 1 5

35 <210> 16
 <211> 25
 <212> PRT
 <213> Artificial

40 <220>
 <223> Secuencia de repetición en héptada

45 <400> 16
 Ile Glu Glu Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Ala
 1 5 10 15

Ala Ile Gln Lys Gln Ile Tyr Arg Met
 20 25

50 <210> 17
 <211> 18
 <212> PRT
 <213> Artificial

ES 2 547 121 T3

<220>

<223> Secuencia de repetición en héptada

5 <400> 17
Ile Glu Glu Ile Gln Lys Gln Ile Ala Ala Ile Gln Lys Gln Ile Tyr
1 5 10 15

Arg Met

<210> 18

<211> 11

10 <212> PRT

<213> Artificial

<220>

<223> Ligador de Gly/Ser con residuos flanqueantes

15 <400> 18
Thr Gly Gly Ser Gly Gly Gly Ser Gly Met Ser
1 5 10

<210> 19

20 <211> 19

<212> PRT

<213> Artificial

<220>

25 <223> Ligador de Gly/Ser con residuos flanqueantes

<400> 19
Thr Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
1 5 10 15

Gly Met Ser

30 <210> 20

<211> 6

<212> PRT

<213> Artificial

35 <220>

<223> Ligador de Gly/Ser

<400> 20
Gly Gly Ser Gly Gly Gly
1 5

40 <210> 21

<211> 10

<212> PRT

<213> Artificial

45 <220>

<223> Ligador de Gly/Ser

<400> 21
Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly
1 5 10

50 <210> 22

<211> 14

ES 2 547 121 T3

<212> PRT
<213> Artificial

<220>
5 <223> Ligador de Gly/Ser

<400> 22
Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly
1 5 10

10 <210> 23
<211> 18
<212> PRT
<213> Artificial

15 <220>
<223> Ligador de Gly/Ser

<400> 23
Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly
1 5 10 15

20 Gly Gly

REIVINDICACIONES

1. Proteína de una sola cadena aislada que consiste en la fórmula HRS1-L1-HRS2-L2-HRS3, en la que HRS1, L1, HRS2, L2 y HRS3 representan fragmentos de secuencia de aminoácidos que están interconectados covalentemente, plegándose espontáneamente dicha proteína en disolución acuosa por medio de los fragmentos HRS1, HRS2 y HRS3 que forman una estructura de superhélice alfa-helicoidal, antiparalela, de triple hebra, y en la que
- 5 cada uno de HRS1, HRS2 y HRS3 es independientemente una secuencia de repetición en héptada que se caracteriza por un patrón de 7 residuos repetido n veces de tipos de aminoácidos, representado como (a-b-c-d-e-f-g)_n o (d-e-f-g-a-b-c)_n y una repetición en héptada parcial C-terminal que termina con un residuo de núcleo de héptada ubicado en una posición a o d, en la que los elementos de patrón "a" a "g" indican posiciones de héptada convencionales en las que se ubican dichos tipos de aminoácidos y n es un número
- 10 igual a o mayor de 2, y
- 15 al menos el 50%, el 70%, el 90% o el 100% de las posiciones de héptada convencionales "a" y "d" están ocupadas por aminoácidos seleccionados del grupo que consiste en valina, isoleucina, leucina, metionina, fenilalanina, tirosina, triptófano, histidina, glutamina, treonina, serina, alanina o derivados de los mismos, en la que al menos el 50%, el 70%, el 90% o en la que el 100% de las posiciones de héptada convencionales
- 20 "b", "c", "e", "f" y "g" están ocupadas por aminoácidos seleccionados del grupo que consiste en glicina, alanina, cisteína, serina, treonina, histidina, asparagina, ácido aspártico, glutamina, ácido glutámico, lisina, arginina o derivados de los mismos, permitiendo la distribución resultante de los tipos de aminoácidos la identificación de dichas secuencias de repetición en héptada, y
- 25 cada uno de L1 y L2 es independientemente un ligador que consiste en de 1 a 30 residuos de aminoácido, incluyendo este ligador cualquier residuo de aminoácido que no puede asignarse de manera inequívoca a una secuencia de repetición en héptada, y
- 30 L1 y L2 tienen una composición de aminoácidos que comprende al menos el 50% de aminoácidos seleccionados del grupo que consiste en glicina, alanina, serina, treonina, prolina o derivados de los mismos.
2. Proteína aislada según la reivindicación 1, en la que al menos el 50%, el 70%, el 90% o en la que el 100% de las posiciones de héptada convencionales "a" y "d" están ocupadas por aminoácidos seleccionados del grupo que consiste en valina, isoleucina, leucina, metionina o derivados de los mismos.
- 35 3. Proteína aislada según cualquiera de las reivindicaciones 1 a 2, en la que al menos el 50%, el 70%, el 90%, o en la que el 100% de las posiciones de héptada convencionales "a" y "d" están ocupadas por isoleucinas.
- 40 4. Proteína aislada según cualquiera de las reivindicaciones 1 a 3, en la que L1 y L2 tienen una composición de aminoácidos que comprende al menos el 50%, el 70%, el 90% o que comprende el 100% de aminoácidos seleccionados del grupo que consiste en glicina, alanina, cisteína, prolina, serina, treonina, histidina, asparagina, ácido aspártico, glutamina, ácido glutámico, lisina, arginina o derivados de los mismos.
- 45 5. Proteína aislada según cualquiera de las reivindicaciones 1 a 4, en la que L1 y L2 tienen una composición de aminoácidos que comprende al menos el 70%, el 90% o que comprende el 100% de aminoácidos seleccionados del grupo que consiste en glicina, alanina, serina, treonina, prolina o derivados de los mismos.
- 50 6. Proteína aislada según cualquiera de las reivindicaciones 1 a 5, en la que L1 y L2 tienen una composición de aminoácidos que comprende al menos el 50%, el 70%, el 90% o que comprende el 100% de aminoácidos glicina y/o serina.
- 55 7. Proteína aislada según cualquiera de las reivindicaciones 1 a 6, en la que el número de residuos de aminoácido de cada uno de L1 y L2 equivale a menos de la mitad del número de residuos de aminoácido de la secuencia de repetición en héptada que precede a los respectivos L1 o L2.
- 60 8. Proteína aislada según cualquiera de las reivindicaciones 1 a 7, en la que residuos de aminoácido cerca de los extremos terminales de L1 y/o L2 estabilizan los extremos alfa-helicoidales de la estructura de superhélice.
9. Proteína aislada según cualquiera de las reivindicaciones 1 a 8, en la que residuos de aminoácido cerca de los extremos terminales de L1 y/o L2 promueven la formación de un giro local en la estructura.
- 65 10. Proteína aislada según cualquiera de las reivindicaciones 1 a 9, en la que las posiciones de héptada

convencionales "e" y "g" están ocupadas por glutaminas.

- 5
11. Proteína aislada según cualquiera de las reivindicaciones 1 a 10, en la que las posiciones de héptada convencionales "b", "c" y "f" son aminoácidos polares que promueven la solubilidad.
12. Proteína aislada según cualquiera de las reivindicaciones 1 a 11, que se pliega en disolución acuosa que tiene un pH de entre 1 y 13, o entre 2 y 12, o entre 3 y 11, o entre 4 y 10, o entre 5 y 9.
- 10
13. Proteína aislada según cualquiera de las reivindicaciones 1 a 12, que se pliega en disolución acuosa que tiene una temperatura de entre 0°C y 100°C, o entre 0°C y 80°C, o entre 0°C y 60°C.
14. Proteína aislada según cualquiera de las reivindicaciones 1 a 13, que se pliega en disolución acuosa que tiene una fuerza iónica de entre 0 y 1,0 molar.
- 15
15. Uso de la proteína aislada según cualquiera de las reivindicaciones 1 a 14 como armazón.
16. Ácido nucleico que codifica para una proteína según cualquiera de las reivindicaciones 1 a 14.
- 20
17. Vector que comprende un ácido nucleico según la reivindicación 16.
18. Célula huésped que comprende un ácido nucleico o vector según la reivindicación 16 ó 18.
- 25
19. Método para la producción de una proteína según cualquiera de las reivindicaciones 1 a 14 que comprende introducir un ácido nucleico o vector en una célula huésped, cultivar dicha célula huésped en un medio en condiciones en las que se expresa el ácido nucleico y se produce la proteína, y aislar la proteína de dicha célula huésped y/o dicho medio.

FIGURA 1

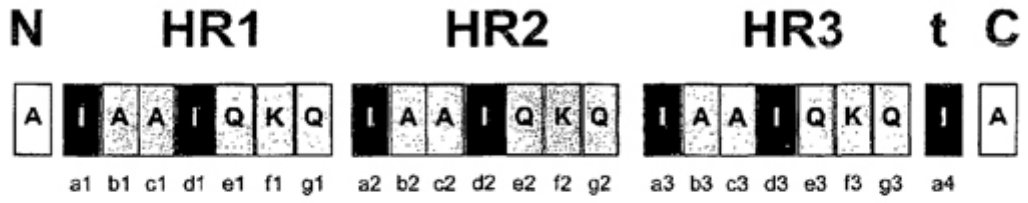


FIGURA 2

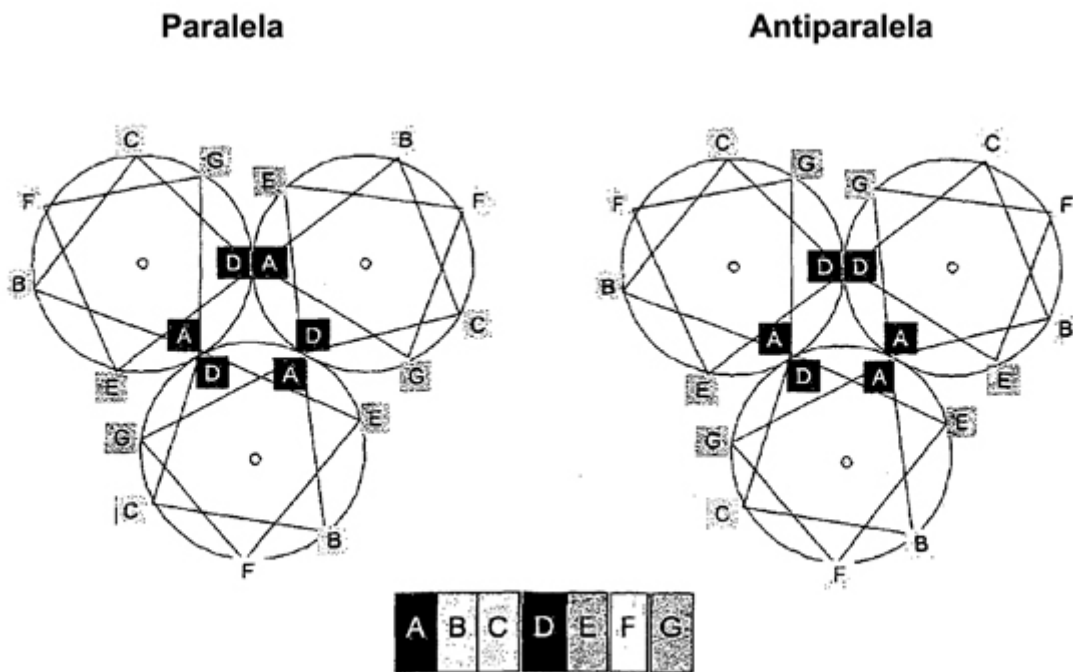


FIGURA 3

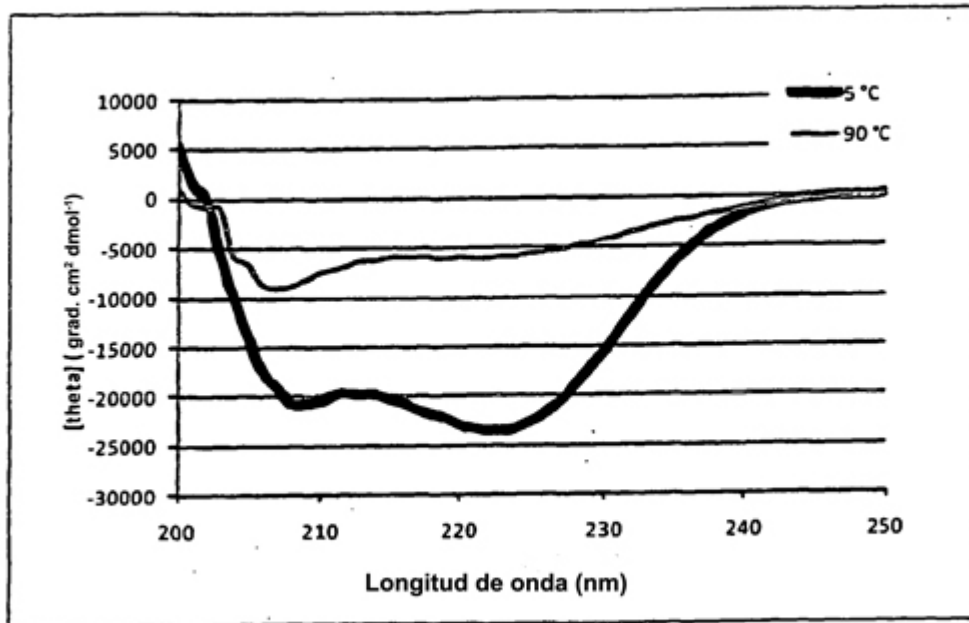


FIGURA 4

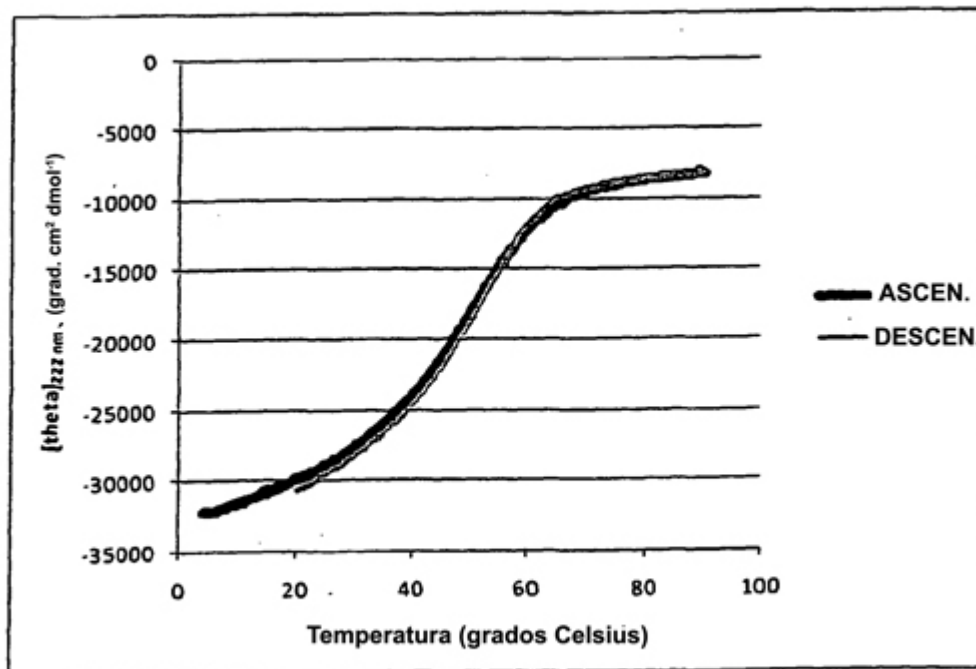


FIGURA 5

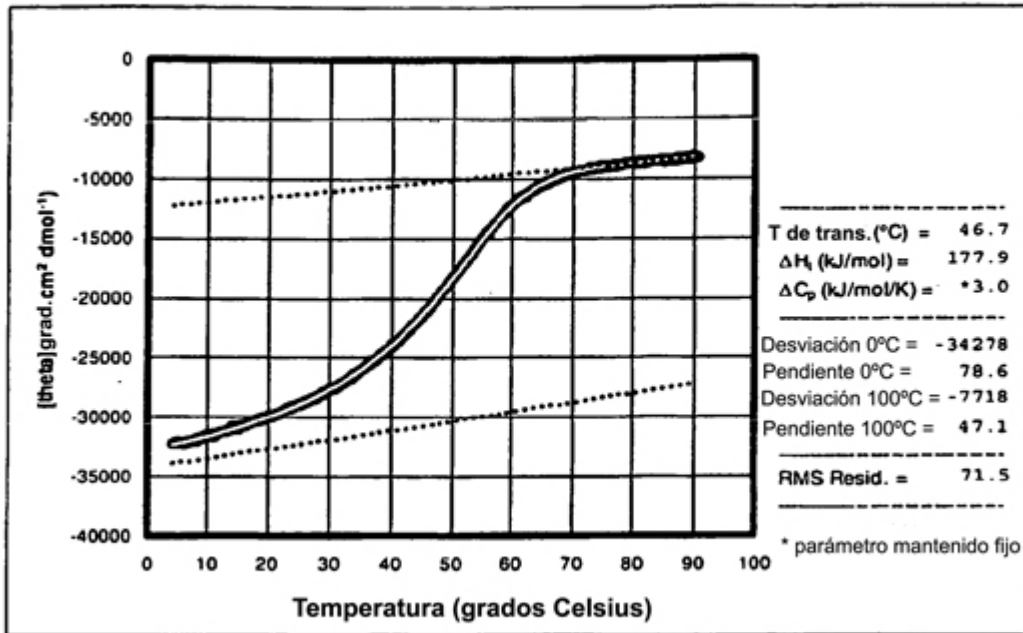


FIGURA 6

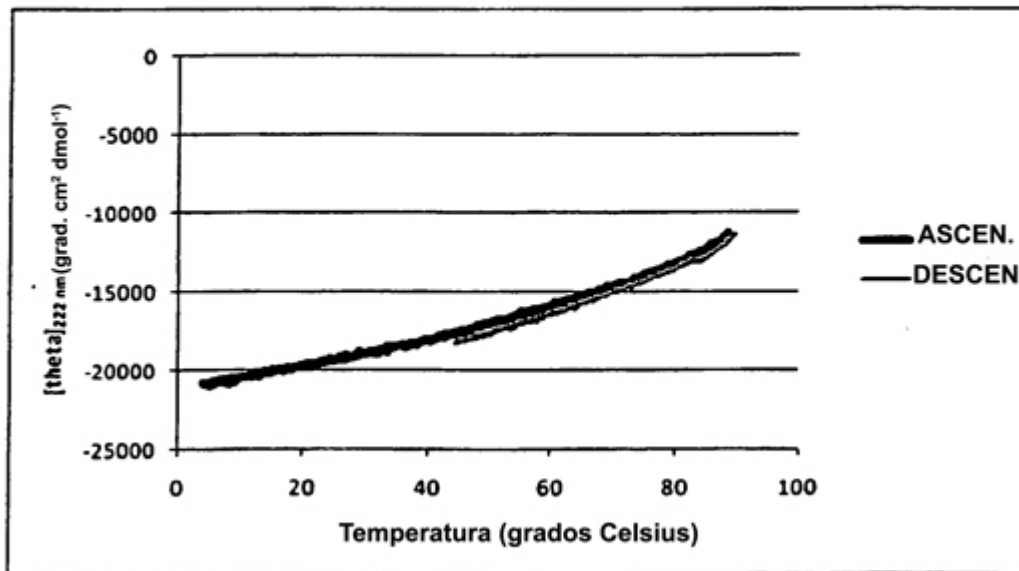


FIGURA 7

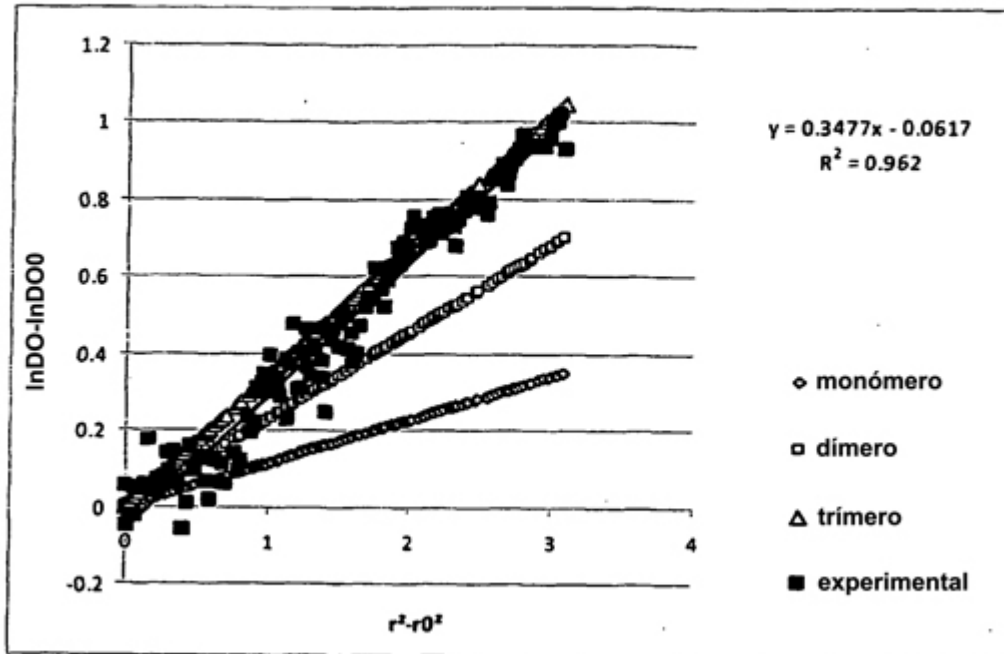


FIGURA 8

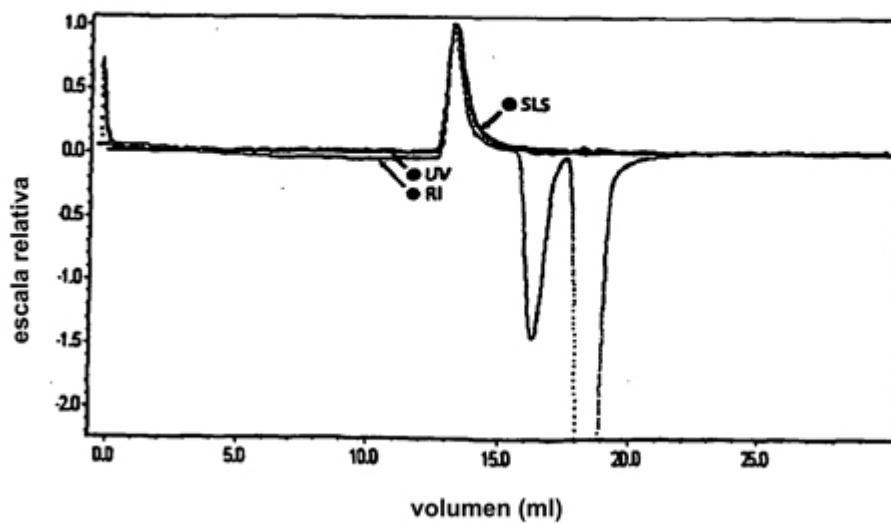


FIGURA 9

scQ2aI_L8 a d a d a d a d	
N	MGHHHHHHHHHHSSGHIEGRHMS
HRS1	IEEIQKQIAAIQKQIAAIQKQIYRM
L1	TGGSGGGSGMS
HRS2	IEEIQKQIAAIQKQIAAIQKQIYRM
L2	TGGSGGGSGMS
HRS3	IEEIQKQIAAIQKQIAAIQKQIYRM
C	TP
Completa	MGHHHHHHHHHHSSGHIEGRHMSIEEIQKQIAAIQKQIAAIQKQIYRMTGGSGGGSG MSIEEIQKQIAAIQKQIAAIQKQIYRMTGGSGGGSGMSIEEIQKQIAAIQKQIAAIQ KQIYRMTTP

scQ2aI_L16 a d a d a d a d	
N	MGHHHHHHHHHHSSGHIEGRHMS
HRS1	IEEIQKQIAAIQKQIAAIQKQIYRM
L1	TGGSGGGSGGGSGGGSGMS
HRS2	IEEIQKQIAAIQKQIAAIQKQIYRM
L2	TGGSGGGSGGGSGGGSGMS
HRS3	IEEIQKQIAAIQKQIAAIQKQIYRM
C	TP
Completa	MGHHHHHHHHHHSSGHIEGRHMSIEEIQKQIAAIQKQIAAIQKQIYRMTGGSGGGSG GGSGGGSGMSIEEIQKQIAAIQKQIAAIQKQIYRMTGGSGGGSGGGSGGGSGMSIEE IQKQIAAIQKQIAAIQKQIYRMTTP

FIGURA 10

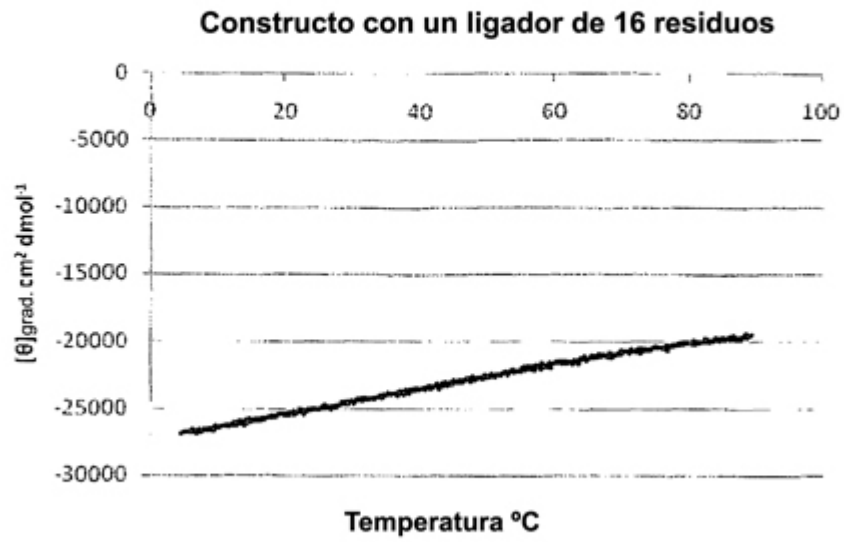


FIGURA 11

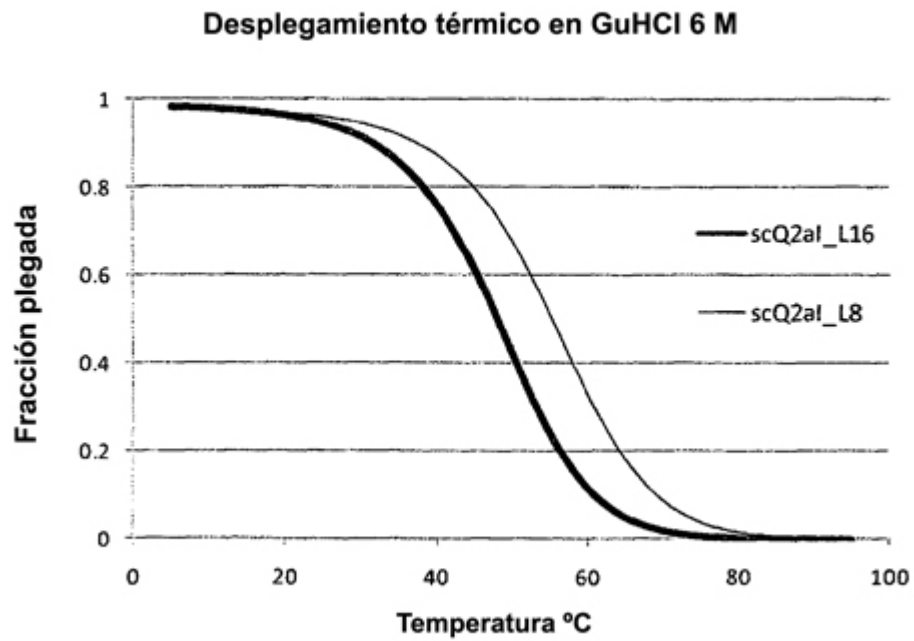


FIGURA 12

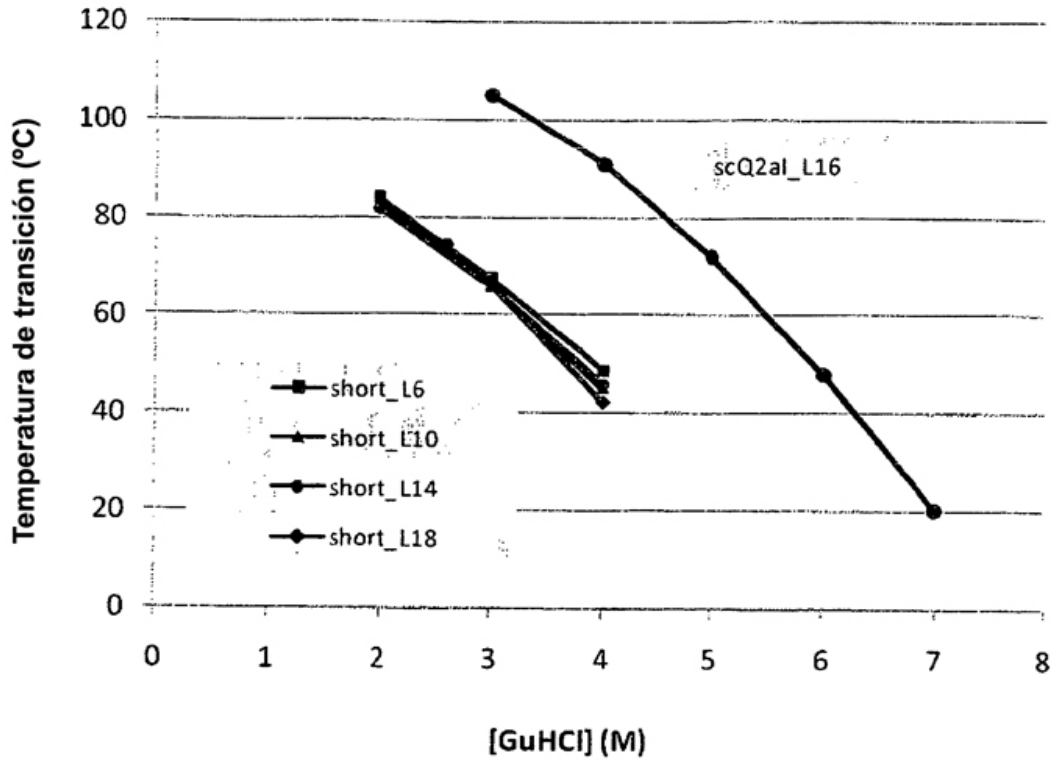


FIGURA 13

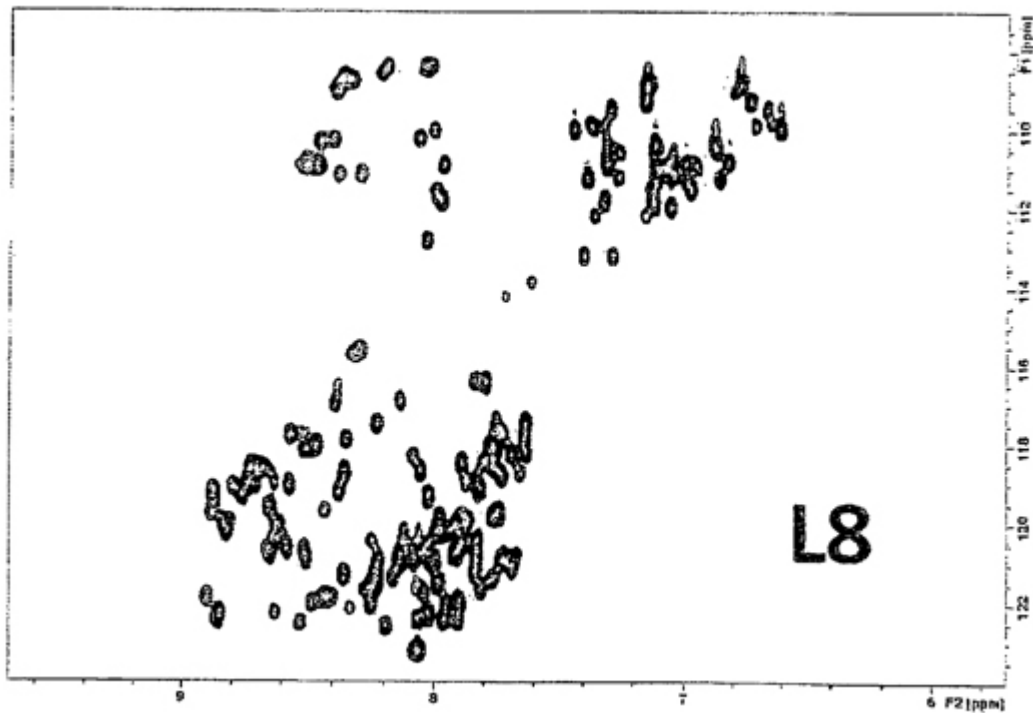
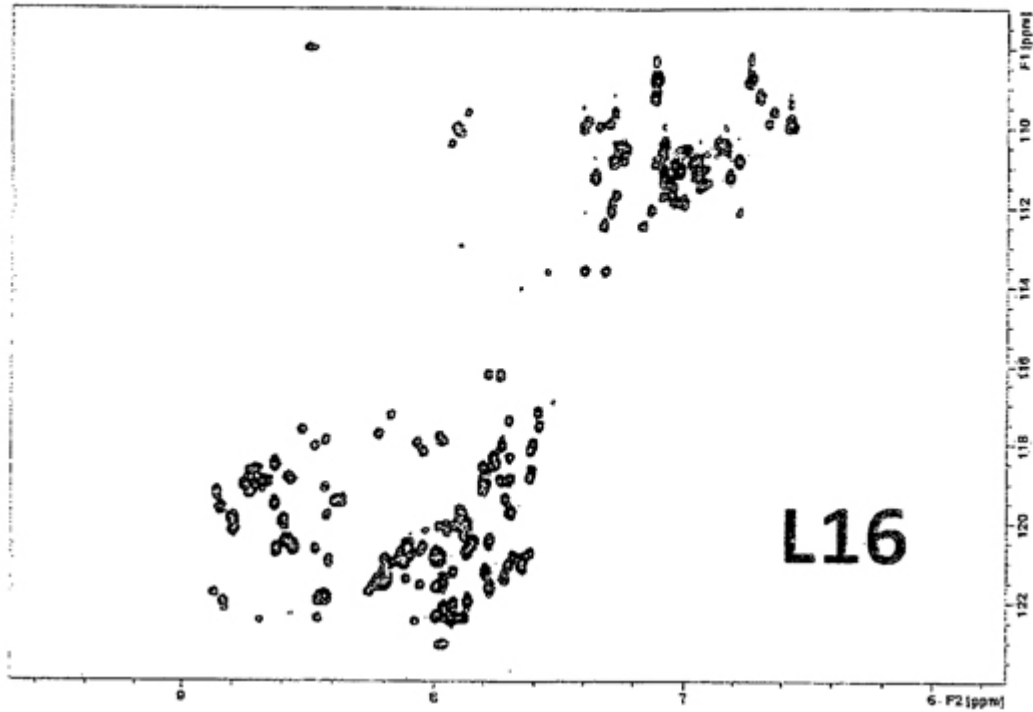


FIGURA 14

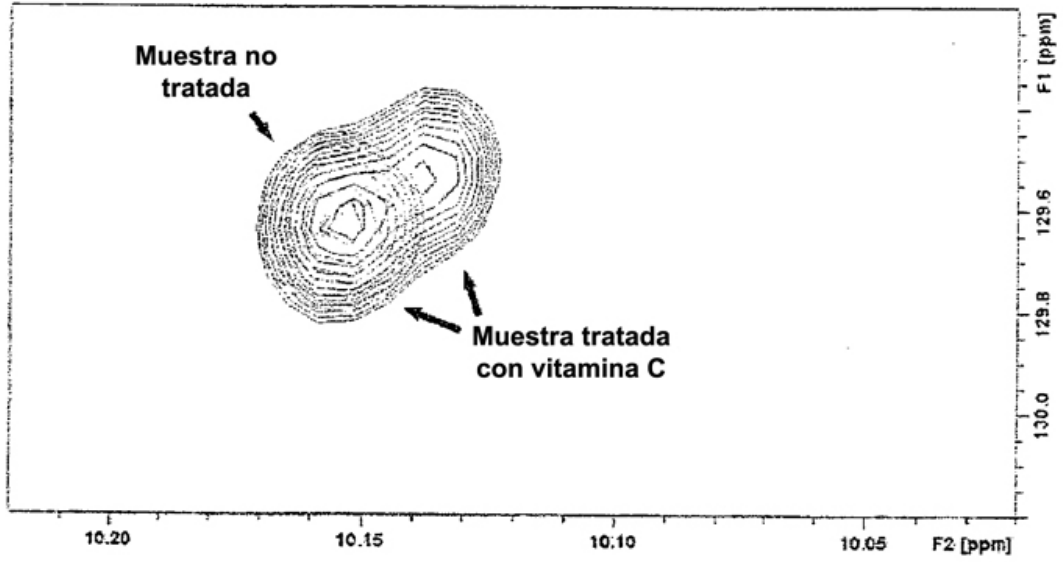


FIGURA 15

