

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 417 781**

51 Int. Cl.:

G06F 19/16 (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.06.2009 E 09767827 (0)**

97 Fecha y número de publicación de la concesión europea: **29.05.2013 EP 2310970**

54 Título: **Procedimientos para identificar regiones de unión a macromolécula y propensas a la agregación en proteínas y usos de los mismos**

30 Prioridad:

20.06.2008 US 74466 P

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

09.08.2013

73 Titular/es:

NOVARTIS AG (50.0%)

Lichtstrasse 35

4056 Basel, CH y

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

(50.0%)

72 Inventor/es:

CHENNAMSETTY, NARESH;

HELK, BERNHARD;

TROUT, BERNHARDT;

KAYSER, VEYSEL y

VOYNOV, VLADIMIR

74 Agente/Representante:

IZQUIERDO FACES, José

ES 2 417 781 T3

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

DESCRIPCIÓN

Procedimientos para identificar regiones de unión a macromolécula y propensas a la agregación en proteínas y usos de los mismos.

5

ANTECEDENTES DE LA INVENCION

El entendimiento y el control de la estabilidad de proteínas ha sido un empeño codiciado para biólogos, químicos e ingenieros. El primer vínculo entre sustitución de aminoácidos y enfermedad (Ingram. *Nature*. 1957, 180(4581):326-8) ofreció una perspectiva nueva y esencial de la estabilidad de las proteína en salud y enfermedad. El reciente enorme aumento de productos farmacéuticos basados en proteínas ha supuesto un nuevo reto. Las proteínas terapéuticas se almacenan en líquido durante varios meses a concentraciones muy altas. El porcentaje de especies no monoméricas aumenta con el tiempo. Como forma agregada, no solo disminuye la eficacia del producto, sino que pueden producirse efectos secundarios tales como respuesta inmunológica tras la administración. El asegurar la estabilidad de los productos farmacéuticos de proteína para la estabilidad en almacén del producto es imprescindible.

Debido a su potencial en la cura de diversas enfermedades, los anticuerpos actualmente constituyen la clase más rápidamente creciente de productos terapéuticos humanos (Carter. *Nature Reviews Immunology*. 2006, 6(5), 343). Desde 2001, su mercado ha ido creciendo a una tasa de crecimiento promedio anual del 35%, la mayor tasa entre todas las categorías de fármacos biotecnológicos (S. Aggarwal, *Nature. BioTech*. 2007, 25 (10) 1097).

Los anticuerpos terapéuticos se preparan y se guardan en disoluciones acuosas a altas concentraciones, según se requiera para el tratamiento de enfermedad. Sin embargo, estos anticuerpos son termodinámicamente inestables bajo estas condiciones y se degradan debido a agregación. La agregación conduce a su vez a una disminución en la actividad del anticuerpo, haciendo el fármaco ineficaz y puede incluso generar una respuesta inmunológica. Como tal, hay una necesidad urgente de desarrollar un entendimiento mecanístico de cómo estos anticuerpos, y de hecho proteínas en general, se agregan, para descubrir qué regiones de la proteína participan en la agregación, y para desarrollar estrategias para impedir la agregación.

Estos efectos son particularmente importantes para agentes terapéuticos de anticuerpos. Un enfoque para la estabilización de anticuerpos es injertar los bucles de CDR que confieren especificidad de unión a antígeno sobre una región estructural más estable (Ewert, Honegger y Pluckthun, *Biochemistry*. 2003, 42(6): 1517-28). Este enfoque solo funcionará si la secuencia de aminoácidos en los bucles de CDR no es la fuerza de agregación motora, y si el injerto de los bucles de CDR sobre una región estructural más estable no cambia la especificidad de unión a antígeno.

La tecnología relacionada con predecir regiones propensas a la agregación de proteínas puede dividirse en dos categorías, 1) modelos fenomenológicos y 2) técnicas de simulación molecular. Los modelos fenomenológicos se basan principalmente en predecir los 'puntos calientes' de agregación de secuencias primarias de proteínas usando propiedades tales como hidrofobia, propensión a hojas β , etc., mientras que las técnicas de simulación molecular usan la estructura tridimensional y la dinámica de proteínas para localizar las regiones propensas a la agregación. La mayoría de las técnicas se han dirigido al entendimiento de la formación de fibrillas amiloides y agregación de otras proteínas pequeñas en las que la formación de hojas β es predominante.

Se han desarrollado modelos fenomenológicos basados en propiedades fisicoquímicas tales como hidrofobia, propensión a hojas β , etc., para predecir las regiones propensas a la agregación de la secuencia primaria de proteínas (Caflisch, *Current Opinion in Chemical Biology*. 2006, 10, 437-444; Chiti y Dobson. *Annu. Rev. Biochem*, 2006, 75: 333-366). Uno de los modelos fenomenológicos iniciales se basó en estudios mutacionales de la cinética de agregación de una proteína globular pequeña 'acilfosfatasa de músculo humano' (AcP) junto con otros péptidos sin estructurar y proteínas nativamente desplegadas (Chiti y col. *Nature*. 2003, 424 pág. 805-808; patente de EE.UU. nº 7379824]. Este estudio reveló simples correlaciones entre la agregación y las propiedades fisicoquímicas tales como propensión a hojas β , hidrofobia y carga. Estos estudios se hicieron en condiciones a las que las proteínas están principalmente sin estructurar. Así, se desarrolló un modelo empírico de tres parámetros que enlaza la secuencia con la propensión a la agregación (Chiti y col. *Nature*. 2003, 424, 805-808). Este modelo también se usó para sugerir variantes de la hormona peptídica de 32 residuos calcitonina para reducir su propensión a la agregación (Fowler y col. *Proc Natl Acad Sci USA*. 2005, 102, 10105-10110). DuBay y colaboradores han extendido la ecuación de tres parámetros (Chiti y col. *Nature*. 2003, 424, 805-808) a una fórmula de siete parámetros que incluye propiedades intrínsecas de la cadena de polipéptidos y factores extrínsecos relacionados con el entorno tal como concentración de péptido, valor de pH y fuerza iónica de la disolución (Dubay y col. *J Mol Biol*. 2004, 341, 1317-1326). Usando este modelo pudieron reproducir las tasas de agregación *in vitro* de un amplio intervalo de péptidos y proteínas sin estructurar. Sin embargo, la principal limitación del modelo de siete parámetros es que a todos los residuos en la secuencia se les dio la misma importancia relativa. Esto no es coherente con observación experimental y de simulaciones que muestra que ciertas regiones son más importantes que otras, dependiendo de sus propensiones de la estructura secundaria. Recientemente, este análisis se extendió adicionalmente para incluir factores de protección para describir la agregación de cadenas de polipéptidos estructuradas (Tartaglia, G. G.,

Pawar, A. P., Campioni, S., Dobson, C. M., Chiti, F. y Vendruscolo, M. *J Mol Biol* (2008) en prensa). Algunos de los sitios predichos estuvieron de acuerdo con los sitios propensos a la agregación conocidos para proteínas tales como lisozima, mioglobina, etc. Se desarrolló un modelo fenomenológico sin parámetros libres (Tartaglia y col. *Protein Sci.* 2004, 13, 1939-1941; Tartaglia y col. *Protein Sci.* 2005, 14, 2723-2734) para predecir cambios en la tasa de alargamiento de la fibrilla de agregados tras la mutación e identificar segmentos propensos a la agregación. Las propiedades fisicoquímicas usadas son el cambio en la propensión a β tras la mutación, el cambio en el número de residuos aromáticos y el cambio en la carga total. Además, la relación de área superficial accesible se tiene en cuenta si las cadenas laterales naturales y mutantes son ambas polares o ambas apolares, mientras que el momento dipolar de la cadena lateral polar se usa en el caso de mutación apolar a polar (o polar a apolar). Este modelo reprodujo la propensión relativa a la agregación de un conjunto de 26 secuencias de heptapéptidos, que se predijo que favorecieron una disposición de hoja β paralela en registro.

El modelo de DuBay y colaboradores (Dubay y col. *J Mol Biol.* 2004, 341, 1317-1326) ha sido modificado con la inclusión de propensión helicoidal α y modelado hidrófobo, y comparando la puntuación de propensión a la agregación de una secuencia de aminoácidos dada con una propensión promedio calculada para un conjunto de secuencias de longitud similar (Pawar y col., *J Mol Biol.* 2005, 350, 379-392). Este modelo se ha validado sobre los segmentos propensos a la agregación de tres cadenas de polipéptidos nativamente desplegados: A β 42, α -sinucleína y la proteína tau.

Se desarrolló otro algoritmo llamado TANGO (Fernandez-Escamilla y col., *Nat Biotechnol.* 2004, 22, 1302-1306), que equilibra los mismos parámetros fisicoquímicos, complementados por la suposición de que un aminoácido está completamente enterrado en el estado agregado. Esto se basa en la propensión de la estructura secundaria y estimación de penalización por desolvatación para predecir regiones de agregación β de una secuencia de proteínas, además de efectos mutacionales. A diferencia de los modelos tratados anteriormente, TANGO tiene en cuenta la estabilidad en estado nativo usando el campo de fuerza FOLD-X. Aunque no es posible calcular tasas absolutas de agregación con TANGO, proporciona una comparación cualitativa entre péptidos o proteínas que se diferencian significativamente en secuencia. Serrano y colaboradores (Linding y col., *J Mol Biol.* 2004, 342, 345-353) han usado TANGO para analizar la propensión a la agregación β de un conjunto de proteínas globulares no redundantes con un límite superior del 40% de identidad de secuencias.

Otro algoritmo, predicción de la agregación de estructuras amiloides, de Prediction of Amiloid StrucTure Aggregation (PASTA), se introdujo recientemente editando una función de energía por emparejamiento para residuos orientados los unos hacia los otros dentro de una hoja β (Trovato y col., *Protein Engineering, Design & Selection.* 2007, 20(10), 521 - 523; Trovato y col., *PLoS Comput.* 2006, 2, 1608-1618; Trovato y col., *J. Phys.: Condens. Matter.* 2007 19, 285221). Yoon y Welsh (Yoon y Welsh, *Protein Sci.* 2004, 13: 2149-2160) han desarrollado un enfoque basado en estructura para detectar la propensión a la agregación β de un segmento de proteína condicionado por el número de contactos terciarios. Usando una ventana de siete residuos de deslizamiento, se sugirió que segmentos con una fuerte tendencia a hoja β en un entorno estrechamente empaquetado (es decir, con un alto número de contactos terciarios) eran el mediador local de la formación de fibrillas.

Aunque se mostró que los modelos fenomenológicos descritos anteriormente cumplían bien para péptidos pequeños y proteínas desnaturalizadas, las propensiones a la agregación podrían diferenciarse para proteínas globulares tales como anticuerpos en los que la estructura terciaria y la estabilidad del estado nativo son muy importantes.

Técnicas de simulación molecular para predecir regiones propensas a la agregación y estudiar el mecanismo de la agregación han empleado generalmente modelos de simulación más simples (Ma y Nussinov. *Curr. Opin. Chem. Biol.* 2006, 10, 445-452; Cellmer y col., *TRENDS in Biotechnology* 2007, 25(6), 254). El menos detallado de los modelos de simulación empleados fue el modelo de red cristalina, en el que cada residuo se representa como una perla que ocupa un único sitio sobre una red cristalina tridimensional. Siguió modelos más detallados, tales como el modelo de resolución intermedia, pero sufrieron la misma incapacidad para representar con exactitud estructuras secundarias y terciarias de la proteína.

A diferencia de modelos más simples, los modelos atomísticos incluyen todos los detalles atomísticos tales como enlaces de hidrógeno y así con más exactos que los modelos de red cristalina o de resolución intermedia. Tales modelos atomísticos se han usado tanto con un disolvente explícito como con un disolvente implícito en el que el disolvente se trata como un continuo. El modelo explícito es más exacto pero también más exigente computacionalmente. Después se desarrolló un protocolo de simulación de dinámica molecular para obtener información estructural sobre la agregación β ordenada de polipéptidos amiloidogénicos (Cecchini y col., *J Mol Biol.* 2006, 357, 1306-1321). Sin embargo, debido a que un procedimiento tal es muy exigente computacionalmente, especialmente para proteínas grandes tales como anticuerpos, no parece haber simulación atomística de anticuerpos completos en la bibliografía. Sin embargo, ha habido simulaciones atomísticas de partes pequeñas del anticuerpo, generalmente para el fragmento Fab (Noon y col., *PNAS.* 2002, 99, 6466; Sinha y Smith-Gill, *Cell Biochemistry and Biophysics.* 2005, 43, 253).

Cellmer y col., *TRENDS in Biotechnology* 2007, 25(6), 254, es un artículo de revisión que resume cómo los avances

en la potencia informática y estudios computacionales referentes al plegamiento de proteínas han facilitado el desarrollo de la simulación molecular como una herramienta para investigar el plegamiento erróneo y la agregación de proteínas. Los autores resaltan los éxitos de estudios de agregación de proteínas llevados a cabo por ordenador, tratan futuras posibilidades para el campo e identifican varios problemas relacionados con la biotecnología que se beneficiarían de la simulación molecular.

De groot y col., BMC Structural Biology, 2005 5(1), 18, describe un estudio en el que los autores explotaron los datos experimentales obtenidos en un sistema *in vivo* usando péptido β -amiloide como modelo para derivar las propensiones a agregación individuales de aminoácidos naturales. Estos datos se usaron para generar perfiles de agregación para polipéptidos relacionados con diferentes enfermedades.

Numerosos enfoques existentes para prevenir la agregación de anticuerpos emplean el uso de aditivos en formulaciones de proteínas. Esto es diferente del enfoque directo descrito en el presente documento, en el que el propio anticuerpo se modifica basándose en las regiones propensas a la agregación predichas de simulaciones moleculares. Aditivos comúnmente usados en la estabilización de anticuerpos son sales de bases que contienen nitrógeno tales como arginina, guanidina o imidazol (documento EP0025275). Otros aditivos adecuados para la estabilización son poliéteres (documento EPA0018609), glicerina, albúmina y sulfato de dextrano (patente de EE.UU. n° 4808705), detergentes y tensioactivos tales como tensioactivos basados en polisorbato (publicación DA2652636 y publicación GB2175906 (solicitud de patente de RU n° GB8514349)), chaperonas tales como GroEL (Mendoza, Biotechnol. Tech. 1991, (10) 535-540), tampón citrato (documento WO9322335) o agentes quelantes (documento WO9115509). Aunque estos aditivos permiten estabilizar proteínas a cierto grado en disolución, sufren ciertas desventajas tales como la necesidad de etapas de procesamiento adicionales para la eliminación del aditivo. Así, se requieren nuevos procedimientos para entender los mecanismos implicados en la agregación de proteínas e identificar las regiones de proteína que median en este fenómeno. Tales procedimientos serían útiles en una variedad de áreas de diagnóstico y terapéuticas, y permitirían que composiciones de proteína tales como anticuerpos terapéuticos se estabilizaran directamente sin el uso de aditivos.

RESUMEN DE LA INVENCION

La presente invención proporciona procedimientos y herramientas computacionales basadas, al menos en parte, en simulaciones informáticas que identifican regiones propensas a la agregación de una proteína. Entonces pueden hacerse sustituciones en estas regiones propensas a la agregación para manipular proteínas con estabilidad potenciada y/o una propensión reducida a la agregación.

Además, la presente invención proporciona procedimientos y herramientas computacionales basados, al menos en parte, en simulaciones informáticas que identifican regiones de unión a macromolécula de una proteína. Entonces pueden hacerse sustituciones y deleciones en estas regiones de unión a macromolécula para manipular proteínas con afinidad de unión alterada por la macromolécula.

En un aspecto, la invención proporciona un procedimiento para calcular la propensión a la agregación espacial (SAP) para un átomo particular en una proteína que comprende (a) identificar uno o más átomos en un modelo estructural que representa la proteína, en el que el uno o más átomos están dentro de una región espacial definida centrada en o próxima al átomo particular; (b) calcular, para el uno o más átomos en la región espacial definida, una relación del área accesible al disolvente (SAA) de los átomos con respecto al SAA de átomos en un residuo idéntico que está completamente expuesto; (c) multiplicar cada relación por la hidrofobia del átomo del uno o más átomos; y (d) sumar los productos de la etapa (c); por lo que la suma es la SAP para el átomo particular.

En una realización relacionada, un procedimiento para calcular la propensión a la agregación espacial (SAP) para un átomo particular en una proteína comprende (a) identificar uno o más residuos de aminoácidos en un modelo estructural que representa la proteína, en el que el uno o más residuos de aminoácidos tienen al menos un átomo dentro de una región espacial definida centrada en o próxima al átomo particular; (b) calcular, para los átomos en la región espacial definida, una relación del área accesible al disolvente (SAA) de los átomos con respecto al SAA de átomos en un residuo idéntico que está completamente expuesto, (c) multiplicar cada relación por la hidrofobia del uno o más residuos de aminoácidos como se ha determinado por una escala de hidrofobia de aminoácidos; y (d) sumar los productos de la etapa (c); por lo que la suma es la SAP para el átomo particular.

Se entiende que en realizaciones particulares la región espacial definida es cualquier volumen o región tridimensional. En realizaciones específicas, la región espacial definida está seleccionada del grupo que comprende una esfera, un cubo, un cilindro, una pirámide y un esferoide elíptico. En algunas realizaciones, la región espacial definida es una región que tiene un volumen equivalente a una esfera con un radio de entre 1-30 Å, o más. En algunas realizaciones, el radio puede ser 50 Å o más. En algunas realizaciones preferidas, el radio de la región espacial definida es 5 Å, o 10 Å.

En una realización preferida, la región espacial definida es una esfera que tiene un radio de entre 1-30 Å. En algunas realizaciones, la esfera está centrada en el átomo particular, mientras que en otras realizaciones la región espacial definida o esfera está centrada en un enlace químico o centrada en un punto en el espacio próximo al átomo sobre

el que se calculará la SAP.

5 En algunas realizaciones, la región espacial definida está centrada en un punto en el espacio dentro de 30 Å desde el átomo particular o en algunas realizaciones preferidas la región espacial definida está centrada en un punto en el espacio dentro de 20 Å, dentro de 10 Å, dentro de 5 Å, dentro de 2 Å, dentro de 1 Å desde el átomo particular.

En algunas realizaciones, el uno o más átomos dentro de la región espacial definida son átomos en una cadena lateral del uno o más aminoácidos.

10 En otras realizaciones, uno o más átomos dentro del radio elegido en un modelo estructural pueden estar, o se requiere que estén, en una cadena lateral de uno o más aminoácidos. Alternativamente, el uno o más átomos dentro del radio elegido en un modelo estructural pueden ser, o se requiere que sean, los átomos de la cadena principal de uno o más aminoácidos.

15 El área accesible al disolvente (SAA) que es parte del cálculo de SAP puede calcularse en algunas realizaciones solo sobre átomos en cadenas laterales de aminoácidos o, en algunas realizaciones, solo sobre átomos de la cadena principal. Los átomos de la cadena principal pueden o pueden no incluir los átomos de hidrógeno unidos.

20 En algunas realizaciones particularmente preferidas, el modelo estructural de proteínas se procesa antes del cálculo de la SAP, por ejemplo, realizando una simulación de dinámica molecular que opcionalmente incluye un disolvente. El disolvente puede ser agua, otro disolvente conocido en la técnica, o el disolvente puede estar ausente. En algunas realizaciones particularmente preferidas, el modelo estructural de proteínas se procesa antes del cálculo de la SAP, por ejemplo, realizando una simulación de Monte Carlo.

25 En otro aspecto, el cálculo de la SAP puede comprender realizar adicionalmente simulaciones de dinámica molecular y promediar los valores de SAP calculados durante múltiples etapas de tiempo en la simulación de dinámica molecular. Por ejemplo, la SAP para el átomo particular puede calcularse realizando una simulación de dinámica molecular antes de la etapa (a) anterior y repitiendo las etapas (a)-(d), realizando cada vez otra simulación de dinámica molecular en una pluralidad de etapas de tiempo, produciendo así múltiples sumas como en la etapa (d), y calcular el promedio de las sumas; por lo que el promedio calculado es la SAP para el átomo particular. En otros ejemplos, una simulación de Monte Carlo puede usarse en su lugar o, o en combinación con una simulación de dinámica molecular.

35 En otras realizaciones, las puntuaciones de SAP pueden sumarse durante múltiples aminoácidos, por ejemplo, sumar durante entre 1 y 50 aminoácidos en una región propensa a la agregación o parche superficial sobre un modelo estructural de proteínas. En una realización particularmente preferida, la SAP se suma durante 1-20 aminoácidos, 1-15 aminoácidos, 1-10 aminoácidos, 1-5 aminoácidos, 1-3 aminoácidos, o la SAP puede sumarse a través de 2 aminoácidos adyacentes. En algunas realizaciones, la suma puede tomarse durante aminoácidos adyacentes que pueden ser secuencialmente adyacentes a lo largo de la secuencia de proteínas o espacialmente en la estructura de proteína.

45 En el que los procedimientos requieren una simulación de dinámica molecular, la simulación puede llevarse a cabo usando un paquete de simulación elegido del grupo que comprende o está constituido por ABINIT, AMBER, Ascalaph, CASTEP, CPMD, CHARMM, DL_POLY, FIREBALL, GROMACS, GROMOS, LAMMPS, MDynaMix, MOLLY, MOSCITO, NAMD, Newton-X, ProtoMol, PWscf, SIESTA, VASP, TINKER, YASARA, ORAC y XMD. En realizaciones particularmente preferidas, el paquete de simulación es el paquete de simulación CHARMM. En otras realizaciones preferidas, el paquete de simulación es el paquete de simulación NAMD.

50 En el que los procedimientos requieren realizar cálculos para uno o más átomos dentro de una cadena lateral, residuo o proteína (por ejemplo, calcular SAA para uno o más átomos), será apreciado por el experto que los cálculos pueden ser para átomos, pares de átomos, combinaciones o grupos de átomos, porciones de átomos, o para cada uno de o todos los átomos en una región espacial, cadena lateral, residuo, proteína, etc. Cuando se realizan cálculos caracterizados en las metodologías de la invención, el experto también apreciará que los cálculos (por ejemplo, cálculos de SAP) también pueden hacerse para residuos de aminoácidos, cadenas laterales y similares, que comprenden átomos, grupos de átomos, etc.

60 En realizaciones adicionalmente preferidas, el modelo estructural es un modelo de estructuras cristalinas de rayos X de la proteína, o porción de la misma; o el modelo estructural puede ser un modelo estructural de proteínas teóricas de la proteína, o porción de la misma. En realizaciones relacionadas, el modelo estructural teórico es un modelo de homología de la proteína o porción de la misma. En otras realizaciones, el modelo estructural teórico es un modelo estructural de proteínas desde el principio de la proteína, o porción de la misma.

65 En otro aspecto, la presente invención proporciona procedimientos para identificar regiones propensas a la agregación sobre una proteína. En una realización, el procedimiento para identificar una región propensa a la agregación sobre una proteína comprende (a) mapear, sobre el modelo estructural, la SAP para átomos en la proteína como se calcula para un átomo particular según cualquier procedimiento descrito en el presente

documento; y (b) identificar una región dentro de la proteína que tiene una pluralidad de átomos que tiene una SAP > 0; en la que la región propensa a la agregación comprende los aminoácidos que comprenden dicha pluralidad de átomos. En algunas realizaciones, el procedimiento puede comprender identificar uno o más aminoácidos que contienen uno o más átomos que tienen una SAP superior a un umbral elegido; en el que la SAP se calcula según cualquier procedimiento descrito en el presente documento y en el que la región propensa a la agregación comprende los aminoácidos identificados.

En otra realización, el procedimiento para identificar una región propensa a la agregación sobre una proteína comprende representar los valores de SAP como se calculan según cualquier procedimiento descrito en el presente documento, adicionalmente calcular picos en la representación del área bajo la curva (ABC) e identificar una o más regiones de proteína con un ABC positiva, en el que la región propensa a la agregación comprende las regiones de proteína identificadas.

En otro aspecto, la invención proporciona procedimientos de preparación de variantes de proteína que presentan una propensión reducida a la agregación. En una realización preferida un procedimiento de preparación de una variante de proteína que presenta una propensión reducida a la agregación comprende sustituir o delecionar al menos un residuo de aminoácido dentro de una región propensa a la agregación en la proteína, en el que la región propensa a la agregación se identifica usando puntuaciones de SAP calculadas según cualquier procedimiento descrito en el presente documento; y en el que, si el residuo de aminoácido se sustituye, se sustituye con un residuo de aminoácido que es más hidrófilo, de forma que se reduce la propensión a la agregación de la variante. En algunas realizaciones, particulares al menos un residuo se sustituye y al menos un residuo se deleciona.

En otra realización, un procedimiento de preparación de una variante de proteína que presenta una propensión reducida a la agregación comprende (a) generar una pluralidad de variantes de proteína sustituyendo en cada variante al menos un residuo dentro de una región propensa a la agregación en la proteína, en el que la región propensa a la agregación se identifica usando puntuaciones de SAP calculadas según cualquier procedimiento descrito en el presente documento, en el que uno o más residuos diferentes, o combinaciones diferentes de residuos, se sustituyen en cada variante; en el que el al menos un residuo se sustituye con un residuo que es más hidrófilo; y (b) seleccionar una variante de proteína preparada como en (a) que presenta una propensión reducida a la agregación.

En algunas realizaciones, el aminoácido que está seleccionado para sustitución es el aminoácido más hidrófobo (como se ha determinado por una escala de hidrofobia reconocida en la materia) en una región propensa a la agregación. En realizaciones específicas, el aminoácido seleccionado para sustitución es Phe, Leu, Ile, Tyr, Trp, Val, Met, Pro, Cys, Ala o Gly. En tales realizaciones específicas, el aminoácido más hidrófilo que está sustituido en la proteína puede seleccionarse del grupo constituido por Thr, Ser, Lys, Gln, Asn, His, Glu, Asp y Arg. Frecuentemente, la escala de hidrofobia preferida para determinar qué residuos son más o menos hidrófilos o hidrófobos que otros es la escala de hidrofobia de Black y Mould.

En algunas realizaciones, al menos dos residuos de aminoácidos dentro de la región propensa a la agregación se sustituyen. En realizaciones relacionadas, al menos tres residuos de aminoácidos dentro de la región propensa a la agregación se sustituyen. Por tanto, en realizaciones similares, al menos un residuo se sustituye con más de una región propensa a la agregación dentro de la proteína.

En realizaciones preferidas, los procedimientos descritos en el presente documento se aplican a una proteína que está seleccionada del grupo constituido por un anticuerpo, un fragmento Fab, un fragmento Fab', un fragmento Fd, un fragmento Fv, un fragmento F(ab')₂ y un fragmento Fc.

En otras realizaciones preferidas, los procedimientos descritos en el presente documento se aplican a una proteína que está seleccionada del grupo constituido por una citocina, una quimiocina, una lipocina, una miocina, un neurotransmisor, una neurotrofina, una interleucina o un interferón. En algunas realizaciones específicas, la proteína puede ser una hormona o factor de crecimiento, un receptor o dominio de receptor, o un neurotransmisor o neurotrofina. En algunas realizaciones, la proteína es un peptidomimético, una proteína modificada, una proteína que comprende aminoácidos no naturales o una proteína que comprende aminoácidos inusuales.

En otro aspecto, la divulgación proporciona procedimientos para calcular el SAA eficaz para un residuo de aminoácido en una proteína. Un procedimiento preferido para calcular el SAA eficaz para un residuo de aminoácido en una proteína comprende (a) calcular para un aminoácido una relación del área accesible al disolvente (SAA) de átomos en el aminoácido con respecto al SAA de átomos en un residuo idéntico que está completamente expuesto; (b) multiplicar la relación por la hidrofobia del aminoácido como se ha determinado por una escala de hidrofobia de aminoácidos; por lo que el producto es el SAA eficaz para el aminoácido. Además, el SAA eficaz para un residuo de aminoácido en una proteína puede calcularse mediante un procedimiento que comprende además sumar el SAA eficaz durante 3 aminoácidos, o en algunas realizaciones 2, 4, 5 ó 6 aminoácidos, que son adyacentes en la secuencia de proteínas.

En otro aspecto, la invención también incluye procedimientos para identificar una región de unión a macromolécula

sobre una proteína que comprende (a) mapear, sobre un modelo estructural de la proteína, la SAP para átomos en la proteína como se calcula para un átomo particular según uno cualquiera de los aspectos precedentes; y (b) identificar una región dentro de la proteína que tiene una pluralidad de átomos que tiene una SAP > 0; en el que la región de unión a macromolécula comprende los aminoácidos que comprenden dicha pluralidad de átomos.

En otro aspecto, la invención incluye procedimientos para identificar una región de unión a macromolécula sobre una proteína, que comprende identificar uno o más aminoácidos que contienen uno o más átomos que tienen una SAP superior a un umbral elegido; en el que la SAP se calcula según el procedimiento de uno cualquiera de los aspectos previos y en el que la región de unión a macromolécula comprende los aminoácidos identificados.

En otro aspecto, la invención incluye procedimientos para identificar una región de unión a macromolécula sobre una proteína, que comprende representar los valores de SAP como se calculan en uno cualquiera de los aspectos precedentes, calcular los picos en la representación, el área bajo la curva (ABC) e identificar una o más regiones de proteína con un ABC positiva, en el que la región de unión a macromolécula comprende las regiones de proteína identificadas.

En otro aspecto, la invención incluye procedimientos de preparación de una variante de proteína que presenta una afinidad de unión reducida por una macromolécula, que comprende sustituir o delecionar al menos un residuo de aminoácido dentro de una región de unión a macromolécula para la macromolécula en la proteína, en el que la región de unión a macromolécula se identifica usando puntuaciones de SAP calculadas según uno cualquiera de los aspectos previos; y en el que, si el residuo de aminoácido se sustituye, se sustituye con un residuo de aminoácido que es más hidrófilo, de forma que se reduce la afinidad de unión por la macromolécula de la variante. En ciertas realizaciones, al menos un residuo se sustituye y al menos un residuo se deleciona. En otro aspecto, la invención también incluye procedimientos de preparación de una variante de proteína que presenta una afinidad de unión alterada por una macromolécula, que comprende (a) generar una pluralidad de variantes de proteína sustituyendo en cada variante al menos un residuo dentro de una región de unión a macromolécula por la macromolécula en la proteína, en el que la región de unión a macromolécula se identifica usando puntuaciones de SAP calculadas según uno cualquiera de los aspectos precedentes, en el que uno o residuos diferentes, o combinaciones diferentes de residuos, se sustituyen en cada variante; y (b) seleccionar una variante de proteína preparada como en (a) que presenta una afinidad de unión alterada por la macromolécula. En ciertas realizaciones, el al menos un residuo de aminoácido dentro de la región de unión a macromolécula es el residuo más hidrófobo en la región de unión a macromolécula. En ciertas realizaciones, el al menos un residuo de aminoácido dentro de una región propensa a la agregación es Phe, Leu, Ile, Tyr, Trp, Val, Met, Pro, Cys, Ala o Gly. En ciertas realizaciones, el residuo de aminoácido que es más hidrófilo está seleccionado del grupo constituido por Thr, Ser, Lys, Gln, Asn, His, Glu, Asp y Arg. En ciertas realizaciones, el residuo de aminoácido que es más hidrófilo es un aminoácido inusual, no natural o modificado. En ciertas realizaciones, el residuo de aminoácido que es más hidrófilo se determina según la escala de hidrofobia de Black y Mould. En ciertas realizaciones, al menos dos residuos de aminoácidos dentro de la región de unión a macromolécula se sustituyen. En ciertas realizaciones, al menos tres residuos de aminoácidos dentro de la región de unión a macromolécula se sustituyen. En ciertas realizaciones, al menos un residuo se sustituye con más de una región propensa a la agregación dentro de la proteína. En ciertas realizaciones, la región propensa a la agregación se identifica según el procedimiento de uno cualquiera de los aspectos precedentes para identificar una región propensa a la agregación sobre una proteína. En ciertas realizaciones, que pueden combinarse con las realizaciones precedentes, la macromolécula es otra proteína, un polinucleótido o un polisacárido. En ciertas realizaciones, que pueden combinarse con las realizaciones precedentes, la proteína está seleccionada del grupo constituido por un anticuerpo, un fragmento Fab, un fragmento Fab', un fragmento Fd, un fragmento Fv, un fragmento F(ab')₂ y un fragmento Fc. En ciertas realizaciones, que pueden combinarse con las realizaciones precedentes, la proteína es una citocina, una quimiocina, una lipocina, una miocina, un neurotransmisor, una neurotrofina, una interleucina o un interferón. En ciertas realizaciones, que pueden combinarse con las realizaciones precedentes, la proteína es una hormona o factor de crecimiento. En ciertas realizaciones, la macromolécula es un receptor de hormona o receptor de factor de crecimiento. En ciertas realizaciones, la proteína es un receptor o dominio de receptor. En ciertas realizaciones, la macromolécula es un agonista de receptor o un antagonista de receptor del receptor o dominio de receptor. En ciertas realizaciones, que pueden combinarse con las realizaciones precedentes, la proteína es un neurotransmisor o neurotrofina. En ciertas realizaciones, la macromolécula es un receptor de neurotransmisor o receptor de neurotrofina.

En otro aspecto, la invención también incluye un procedimiento para la preparación de una composición farmacéutica que comprende una variante de proteína que presenta una propensión alterada para interacción con un componente de unión, que comprende formular una variante de proteína obtenida según un procedimiento de cualquiera de los aspectos precedentes junto con un vehículo, adyuvante y/o excipiente farmacéuticamente aceptable.

DESCRIPCIÓN DETALLADA DE LA INVENCION

La presente invención trata la necesidad sin satisfacer de entender más profundamente el mecanismo de agregación de proteínas, y de identificar las regiones de proteína implicadas en la agregación. La invención proporciona, al menos en parte, una tecnología de simulación que puede usarse, simultáneamente con los procedimientos

experimentales descritos en el presente documento, para mejorar la estabilidad de posiblemente todas las proteínas terapéuticas contra la agregación. Esta tecnología presenta un enorme potencial científico y comercial considerando que las terapias basadas en anticuerpos están creciendo al mayor ritmo de entre todas las clases de agentes terapéuticos humanos. La agregación es un problema común encontrado en la mayoría de las etapas del desarrollo de fármacos de anticuerpo que impide la rápida comercialización de posibles candidatos a fármacos de anticuerpo. Así, la prevención de la agregación usando los procedimientos descritos en el presente documento podría tener un impacto significativo sobre el desarrollo de fármacos de proteína.

Además, la presente invención trata la necesidad sin satisfacer de identificar con exactitud las regiones de proteína implicadas en la unión con otras macromoléculas cuya unión está frecuentemente mediada, al menos en parte, por grandes parches hidrófobos que pueden identificarse fácilmente usando los procedimientos desvelados en el presente documento. La invención proporciona, al menos en parte, una tecnología de simulación que puede usarse, simultáneamente con los procedimientos experimentales descritos en el presente documento, para alterar la afinidad de unión de posiblemente todas las interacciones moleculares con proteínas que están mediadas, al menos en parte, por grandes parches hidrófobos. Esta tecnología presenta un enorme potencial científico y comercial considerando que las terapias basadas en proteínas están creciendo al mayor ritmo de entre todas las clases de agentes terapéuticos humanos. La capacidad para alterar una afinidad de unión a agente terapéutico de proteína por una o más macromoléculas puede usarse para mejorar la eficacia y reducir o eliminar actividades mediadas por una región de unión a macromolécula secundaria no deseada.

La presente invención proporciona, entre otros, procedimientos para reducir o prevenir la agregación de una proteína o alterar la afinidad de unión por una macromolécula. En particular, se proporcionan procedimientos para identificar regiones hidrófobas sobre una estructura de proteína que pueden participar en interacciones de proteínas, interacciones proteína-macromolécula o agregación de proteínas. Los procedimientos proporcionados se basan en una nueva técnica desvelada en el presente documento como la "propensión a la agregación espacial" o "SAP". La herramienta de SAP también identifica correctamente las regiones del anticuerpo propensas a unirse con otras proteínas. Además de anticuerpos, esta herramienta podría aplicarse ampliamente a todas las proteínas para identificación de las regiones propensas a la agregación o las regiones que se unen a otras proteínas o ligandos. Los procedimientos de la presente invención pueden aplicarse a cualquier proteína para la que esté disponible una estructura tridimensional o para la que puede crearse una estructura tridimensional usando modelado por homología, modelado molecular o determinación estructural desde el principio. En general, la "SAP" puede calcularse de múltiples formas, usando las ecuaciones y metodología descritas en el presente documento, por ejemplo, la SAP puede calcularse sobre un modelo estructural de proteína o puede calcularse como un promedio sobre múltiples etapas de tiempo de una simulación de dinámica molecular de un modelo estructural. Aunque el procedimiento específico de cálculo, y los resultados obtenidos, pueden variar como se describe en el presente documento, el principio subyacente se basa en el hecho de que la SAP es una medida que no solo explica la hidrofobia de residuos en una proteína, sino también la estructura tridimensional de la proteína, y la proximidad de residuos de aminoácidos en la estructura de proteína plegada.

Por "proteína" se indica cualquier secuencia de dos o más aminoácidos (también denominada en el presente documento "residuos de aminoácidos" o "residuos") unidos juntos por enlaces peptídicos entre grupos carboxilo y amino de aminoácidos adyacentes, independientemente de la longitud, modificación post-traducciona, modificación química o función. "Polipéptido", "péptido" y "proteína" se usan indistintamente en el presente documento. En realizaciones preferidas, los procedimientos de la presente invención se aplican a una proteína que es de longitud suficiente para plegarse en una estructura tridimensional. En algunas realizaciones, la proteína es una proteína que se produce naturalmente. En algunas realizaciones, la proteína se sintetiza químicamente. En algunas realizaciones, la proteína es una proteína recombinante, por ejemplo, una proteína híbrida o quimérica. En algunas realizaciones, la proteína es una proteína complejada (por ejemplo, proteínas de interacción complejadas). Las proteínas pueden aislarse (por ejemplo, de una fuente natural o medio químico). En algunas realizaciones, la proteína puede ser una proteína modificada o un peptidomimético. En algunas realizaciones, la proteína puede ser una proteína derivatizada, por ejemplo, una proteína químicamente conjugada (que incluye, pero no se limita a, proteínas conjugadas con polímero (por ejemplo, proteínas PEGiladas). Como se usa en el presente documento, el término "proteína" también está previsto que incluya fragmentos de proteína. Proteínas a modo de ejemplo incluyen anticuerpos (que incluyen, pero no se limitan a, fragmentos, variantes y derivados de los mismos).

De hecho, se concibe que los procedimientos de la presente invención pueden aplicarse a cualquier molécula basada en aminoácidos para la que esté disponible o pueda generarse un modelo estructural. Por ejemplo, los procedimientos descritos en el presente documento pueden aplicarse a proteínas modificadas, o proteínas que incorporan aminoácidos inusuales o no naturales como se describe en el presente documento. En algunas realizaciones, las estructuras de aminoácidos inusuales, no naturales o modificados pueden sustituirse o insertarse computacionalmente en un modelo estructural para la aplicación de los procedimientos descritos en el presente documento. Procedimientos de diseñar experimentalmente análogos de péptidos, derivados y miméticos se conocen en la técnica. Por ejemplo, véase Farmer, P.S. en *Drug Design* (E.J. Ariens, ed.) Academic Press, Nueva York, 1980, vol. 10, pág. 119-143; Ball, J.B. y Alewood, P.F. (1990) *J. Mol. Recognition* 3:55; Morgan, B.A. y Gainor, J.A. (1989) *Ann. Rep. Med. Chem.* 24:243; y Freidinger, R.M. (1989) *Trends Pharmacol. Sci.*, 10:270. Véase también Sawyer, T.K. (1995) "*Peptidomimetic Design and Chemical Approaches to Peptide Metabolism*" en Taylor, M.D. y Amidon,

G.L. (eds.) *Peptide-Based Drug Design: Controlling Transport and Metabolism*, Capítulo 17; Smith, A.B. 3ª y col. (1995) *J. Am. Chem. Soc.* 117:11113-11123; Smith, A.B. 3ª y col. (1994) *J. Am. Chem. Soc.* 116:9947-9962; y Hirschman, R. y col. (1993) *J. Am. Chem. Soc.* 115:12550-12568.

5 Se conoce en la técnica un gran número y variedad de agentes terapéuticos de péptido, polipéptido y proteína, y se espera que se beneficien de los procedimientos de la presente invención. Estos agentes terapéuticos comprenden varias clases muy amplias que incluyen hormonas, proteínas, antígenos, inmunoglobulinas, represores/activadores, enzimas, citocinas, quimiocinas, miocinas, lipocinas, factores de crecimiento, receptores, dominios de receptores, neurotransmisores, neurotrofinas, interleucinas e interferones, entre otros.

10 Hormonas adecuadas que pueden emplearse dentro del alcance de la presente invención incluyen hormonas de proteína, tales como insulina y glucagón que regulan el azúcar en sangre. Como será apreciado por un experto habitual en la materia, las hormonas anotadas se emplean normalmente para el tratamiento de diversas afecciones y enfermedades que incluyen cáncer, enfermedades metabólicas, enfermedad cardiovascular, afecciones hipofisarias y menopausia.

15 Inicialmente se creyó que solo algunas proteínas formaban fibrillas o agregados. Más recientemente se demuestra que muchas más proteínas de las esperadas tienen regiones propensas a la agregación (Fandrich, M., Fletcher, M. A. y Dobson, C. M. (2001) *Nature* 410, 165-166). De hecho, está documentado que péptidos de tan solo 4 residuos pueden formar fibrillas (*J. Biol Chem.*, vol. 277, edición 45, 43243-43246, 8 de noviembre de 2002).

20 Los agentes terapéuticos de proteína representan una cuota creciente del mercado terapéutico. Por ejemplo, la insulina y los glucagones son agentes terapéuticos de proteína importantes que regulan el azúcar en sangre y pueden beneficiarse de los procedimientos descritos en el presente documento. El polipéptido amiloide de los islotes (IAPP) es otra hormona secretada por el páncreas que se usa en el tratamiento de diabetes. Otra proteína de interés es el factor estimulante de colonias de granulocitos, o G-CSF, que es un factor de crecimiento de la sangre que puede usarse para aumentar la producción de glóbulos sanguíneos. El activador tisular del plasminógeno es un destructor de coágulos usado en el tratamiento de accidente cerebrovascular o infarto de miocardio. Además, la eritropoyetina es una hormona producida por el riñón que puede usarse en el tratamiento de SIDA, anemia, insuficiencia renal y otras afecciones. Finalmente, la calcitonina es un péptido que se ha encontrado que es eficaz en el tratamiento de hipercalcemia, enfermedad de Paget y ciertos tipos de osteoporosis.

25 Otros ejemplos de proteínas que se espera que se beneficien de los procedimientos descritos en el presente documento incluyen, sin limitación, ACTH, amilina, angiotensina, angiogenina, péptidos antiinflamatorios, BNP, endorfinas, endotelina, GLIP, factor liberador de hormona de crecimiento (GRF), hirudina, insulintropina, neuropéptido Y, PTH, VIP, hormona liberadora de hormona de crecimiento (GHRH), octreotida, hormonas hipofisarias (por ejemplo, hGH), ANF, factores de crecimiento, bMSH, somatostatina, factor liberador de factor de crecimiento derivado de plaquetas, gonadotropina coriónica humana, hirulog, interferón alfa, interferón beta, interferón gamma, interleucinas, factor estimulante de colonias de granulocitos y macrófagos (GM-CSF), factor estimulante de colonias de granulocitos (G-CSF), menotropinas (urofolitropina (FSH) y LH)), estreptocinasa, urocina, ANF, ANP, inhibidores de la eliminación de ANP, agonistas de hormonas antidiuréticas, péptido relacionado con el gen calcitonina (CGRP), IGF-1, pentigéptido, proteína C, proteína S, timosina alfa-1, análogos de antagonistas de vasopresina, TNF- α negativo dominante, alfa-MSH, VEGF, PYY, y polipéptidos, fragmentos, análogos de polipéptido y derivados de los anteriores.

35 En realizaciones particularmente preferidas, la proteína es un anticuerpo o inmunoglobulina. El término "anticuerpo" se usa en el sentido más amplio y cubre específicamente anticuerpos monoclonales (incluyendo anticuerpos monoclonales de longitud completa), anticuerpos policlonales, anticuerpos multiespecíficos (por ejemplo, anticuerpos biespecíficos), anticuerpos monocatenarios, anticuerpos quiméricos, anticuerpos recombinantes y fragmentos de anticuerpos. Un anticuerpo de longitud completa es una glucoproteína que comprende al menos dos cadenas pesadas (H) y dos cadenas ligeras (L) interconectadas por enlaces disulfuro. El residuo de Asn-297 en C_{H2} está N-glucosilado. Cada cadena pesada comprende una región variable de la cadena pesada (abreviada en el presente documento V_H) y una región constante de la cadena pesada. La región constante de la cadena pesada comprende tres dominios, C_{H1}, C_{H2} y C_{H3}. Los receptores de Fc se unen a la región bisagra inferior de C_{H2} y median en funciones efectoras tales como citotoxicidad mediada por célula dependiente de anticuerpo (ADCC). La proteína A se une en el empalme C_{H2}-C_{H3} de Fc y se usa ampliamente en la purificación de anticuerpos completos. Cada cadena ligera comprende una región variable de la cadena ligera (abreviada en el presente documento V_L) y una región constante de la cadena ligera. La región constante de la cadena ligera comprende un dominio, C_L. Las regiones V_H y V_L pueden subdividirse adicionalmente en regiones de hipervariabilidad, llamadas regiones determinantes de la complementariedad (CDR), intercaladas con regiones que son más conservadas, llamadas regiones estructurales (FR). Cada V_H y V_L está compuesta por tres CDR y cuatro FR, dispuestas del extremo amino al extremo carboxi en el siguiente orden: FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4. Las regiones variables de las cadenas pesadas y ligeras contienen un dominio de unión que interacciona con un antígeno. Así, el término "anticuerpo" englobaría los diversos isotipos o subclases de anticuerpos, por ejemplo, IgA, IgD, IgE, IgG e IgM, o IgG1, IgG2, IgG3 e IgG4. Adicionalmente se incluyen un fragmento Fab, un fragmento monovalente constituido por los dominios V_L, V_H, C_L y C_{H1}; un fragmento F(ab')₂, un fragmento bivalente que comprende dos fragmentos Fab,

ligados por un puente disulfuro en la región bisagra; un fragmento Fab', que es esencialmente un Fab con parte de la región bisagra (véase, FUNDAMENTAL IMMUNOLOGY (Paul ed., 3ª ed. 1993); un fragmento Fd constituido por los dominios V_H y C_{H1}; un fragmento Fv constituido por los dominios V_L y V_H de un único brazo de un anticuerpo, un fragmento dAb (Ward y col., (1989) *Nature* 341:544-546), que consiste en un dominio V_H; una región determinante de la complementariedad (CDR) aislada; y un nanocuerpo, una región variable de la cadena pesada que contiene un único dominio variable y dos dominios constantes.

Como se usa en el presente documento, un "modelo estructural" de proteína es una representación de una estructura secundaria, terciaria y/o cuaternaria tridimensional de la proteína. Un modelo estructural engloba estructuras cristalinas de rayos X, estructuras de RMN, estructuras de proteínas teóricas, estructuras creadas a partir de modelado por homología, modelos de tomografía de proteínas y modelos atomísticos construidos a partir de estudios de microscopía electrónica. Normalmente, un "modelo estructural" no englobará simplemente la secuencia de aminoácidos primaria de una proteína, sino que proporcionará coordenadas para los átomos en una proteína en el espacio tridimensional, mostrando así los pliegues de proteína y posiciones de residuos de aminoácido. En realizaciones preferidas, el modelo estructural analizado es una estructura cristalina de rayos X, por ejemplo, una estructura obtenida de la base de datos de proteínas (PDB, rcsb.org/pdb/home/home.do) o un modelo de homología basado en una estructura conocida de una proteína similar. En realizaciones preferidas, el modelo estructural se reprocesará antes de aplicar los procedimientos de la presente invención. Por ejemplo, el modelo estructural puede pasarse a una simulación de dinámica molecular para permitir que las cadenas laterales de proteína alcancen una conformación más natural, o puede permitirse que el modelo estructural interaccione con disolvente, por ejemplo, agua, en una simulación de dinámica molecular. El pre-procesamiento no se limita a la simulación de dinámica molecular y puede llevarse a cabo usando cualquier medio reconocido en la técnica para determinar el movimiento de una proteína en disolución. Una técnica de simulación alternativa a modo de ejemplo es la simulación de Monte Carlo. Las simulaciones pueden realizarse usando paquetes de simulación o cualquier otro medio de computación aceptable. En ciertas realizaciones, las simulaciones para buscar, sondar o muestrear espacio conformacional de proteínas puede realizarse en un modelo estructural para determinar el movimiento de la proteína.

Una "estructura de proteínas teóricas" es un modelo estructural de proteínas tridimensionales que se crea usando procedimientos computacionales frecuentemente sin ninguna medición experimental directa de la estructura nativa de la proteína. Una "estructura de proteínas teóricas" engloba modelos estructurales creados por procedimientos desde el principio y modelado por homología. Un "modelo de homología" es un modelo estructural de proteínas tridimensionales que es creado por modelado por homología, que normalmente implica comparar una secuencia primaria de la proteína con la estructura tridimensional conocida de una proteína similar. El modelado por homología es muy conocido en la técnica y se describe en Kolinski y col. *Proteins*. 1999; 37(4):592-610; Rost y col.. *B, Protein Sci.* 1996; 5(8):1704-1718, y las patentes de EE.UU. nº 7212924; 6256647; y 6125331. En particular, Xiang (*Curr Protein Pept Sci.* Junio de 2006; 7(3):217-27) proporciona una excelente descripción y revisión de técnicas de modelado por homología que pueden usarse para generar estructuras útiles para los procedimientos de la presente invención. De hecho, cualquier software de modelado por homología conocido en la técnica puede usarse según los presentes procedimientos, por ejemplo, MODELLER (Eswar y col., *Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., Suplemento 15, 5.6.1-5.6.30, 200), SEGMOD/ENCAD (Levitt M. *JMolBiol* 1992; 226:507-533), SWISS-MODEL (Schwede T, Kopp J, Guex N, Peitsch MC. *Nucleic Acids Research* 2003; 31:3381-3385), 3D-JIGSAW (Bates y col., *Proteins: Structure, Function and Genetics, Suppl* 2001; 5:39-46), NEST (Xiang. *Curr Protein Pept Sci.* Junio de 2006; 7(3): 217-227) y BUILDER (Koehl y Delarue. *Curr Opin Struct Biol* 1996; 6(2):222-226.). Para anticuerpos en particular, la estructura de regiones variables de anticuerpos puede obtenerse con exactitud usando el procedimiento de estructuras canónicas (Chothia C y Lesk AM, *J. Mol. Biol.* 1987, 196, 901; Chothia C y col., *Nature* 1989, 342, 877).

En realizaciones particulares, el modelado por homología puede usarse para ensamblar proteínas completas de fragmentos de estructura conocida tales como cuando un fragmento Fab de anticuerpo se modela sobre un fragmento Fc, o cuando un fragmento Fab se crea como una estructura de proteínas teóricas y se modela sobre una estructura cristalina del fragmento Fc. Un experto entenderá que existen diversas posibilidades. En una realización particular, un fragmento Fab puede modelarse sobre diversas estructuras de anticuerpos Fc de diferentes clases o isotipos.

También pueden emplearse modelos desde el principio en los procedimientos de la presente invención. Un "modelo estructural de proteínas desde el principio" es un modelo estructural de proteínas que es creado directamente a partir de la secuencia primaria de proteínas simulando el procedimiento de plegamiento de proteínas usando las ecuaciones conocidas en la química física (Bonneau y Baker. *Annual Review of Biophysics and Biomolecular Structure*. 2001, vol. 30, páginas 173-189; Lesk *Protein* 1997; 1:151-166. *Suppl*; Zemla y col.. *Proteins* 1997; 1:140-150. *Suppl*; Ingwall y col. *Biopolymers* 1968; 6:331-368; y patentes de EE.UU. nº 6832162; 5878373; 5436850; 6512981; 7158891; 6377893; y solicitudes de patente de EE.UU. nº 9/788.006; 11/890.863; y 10/113.219). Normalmente, estructuras experimentalmente determinadas (por ejemplo, estructuras cristalinas de rayos X) y modelos de homología son preferibles para modelos desde el principio, ya que la dificultad en simular el plegamiento de proteínas *de novo* puede conducir, en algunos casos, a modelos estructurales de proteínas imprecisos.

Se entiende que cualquier procedimiento conocido en la técnica para generar una estructura de proteínas teóricas

puede ser útil según la presente invención. Además de los procedimientos descritos anteriormente, en la presente metodología pueden usarse procedimientos tales como aquellos descritos en la reunión Evaluación crítica de técnicas para la predicción de estructuras de proteínas (CASP). Se describen diversos ejemplos en los procedimientos de CASP, por ejemplo, en las publicaciones relacionadas con el 7º Experimento de comunidad amplia sobre la Evaluación crítica de técnicas para la predicción de estructuras de proteínas, Centro de conferencias Asilomar, Pacific Grove, CA, 26-30 de noviembre de 2006 y también en los procedimientos de CASP6. Proteins: Structure, Function, and Bioinformatics, 2005, 61(S7):1-236; procedimientos de CASP5. Proteins: Structure, Function, and Genetics. 2003, 53(S6):333-595; procedimientos de CASP4. Proteins: Structure, Function, and Genetics. 2001, 45(S5):1-199; procedimientos de CASP3. Proteins: Structure, Function, and Genetics, 1999, 37(S3):1-237 (1999)

La presente invención también proporciona un procedimiento de preparación de una variante de proteína que presenta una propensión reducida a la agregación. Como se usa en el presente documento, una "propensión a la agregación" es la tendencia de una proteína para formar agrupaciones o masas. Tales agrupaciones o masas pueden contener dos, o más frecuentemente 3, o más proteínas, normalmente del mismo tipo. Por consiguiente, una proteína que presenta una "propensión reducida a la agregación" es una que, cuando se modifica o trata, forma menos agregados o agregados más pequeños con respecto a la misma proteína que está sin modificar o sin tratar.

El término "inhibir" pretende expresar una reducción medible de un fenómeno, frecuentemente usado en el presente documento en referencia a interacciones o agregación de unión a proteína.

Residuos de aminoácidos, agrupaciones de residuos, regiones de proteína, péptidos, o parches sobre una superficie de proteína, pueden describirse frecuentemente en el presente documento como hidrófilos o hidrófobos. Según los procedimientos de la invención, la propensión a la agregación espacial describe hidrofobia y se calcula, en parte, usando una escala de hidrofobia de aminoácidos conocidos en la técnica. En una realización preferida, la escala de hidrofobia de aminoácidos es la escala expuesta en Black y Mould, Anal. Biochem. 1991, 193, 72-82. En general, según Black y Mould, la hidrofobia de los aminoácidos progresa del siguiente modo (empezando con los residuos más hidrófobos): Phe > Leu = Ile > Tyr ≈ Trp > Val > Met > Pro > Cys > Ala > Gly > Thr > Ser > Lys > Gln > Asn > His > Glu > Asp > Arg. Los valores estandarizados para la hidrofobia, como se informa por Black y Mould, se muestran en la siguiente Tabla 1.

Tabla 1

Ala	0,616
Cys	0,68
Asp	0,028
Glu	0,043
Phe	1
Gly	0,501
His	0,165
Ile	0,943
Lys	0,283
Leu	0,943
Met	0,738
Asn	0,236
Pro	0,711
Gln	0,251
Arg	0
Ser	0,359
Thr	0,45
Val	0,825
Trp	0,878
Tyr	0,88
Asx	0,132
Glx	0,147

Por consiguiente, cuando un aminoácido se selecciona para sustitución mediante los procedimientos de la invención (por ejemplo, por tener una alta puntuación de SAP o por ser identificado que reside en una región propensa a la agregación), se sustituirá con otro aminoácido que es menor en una escala de hidrofobia. Por ejemplo, si el aminoácido metionina se selecciona para sustitución, puede sustituirse con cualquier aminoácido que sea menos hidrófobo, por ejemplo, Pro, Cys, Ala, Gly, etc. En realizaciones particularmente preferidas, un aminoácido hidrófobo se sustituye con Lys. En realizaciones adicionalmente preferidas, un aminoácido hidrófobo se sustituye con Glu, Gln, Asp, Thr o Ser. Por tanto, si un residuo se describe como "más hidrófobo", "más hidrófilo", "el más hidrófobo" o "el más hidrófilo", la determinación de la hidrofobia/hidrofilia se hace según cualquier escala de hidrofobia conocida en la técnica, por ejemplo, la escala preferida de Black y Mould.

En la práctica, cualquier escala reconocida en la técnica de la hidrofobia de aminoácidos puede emplearse mediante los procedimientos de la presente invención. Así, aunque la escala descrita en la Tabla 1 puede usarse durante el cálculo de la propensión a la agregación espacial, otras escalas conocidas en la técnica pueden estar sustituidas. La reciente revisión por Biswas y col. (*J. Chromatogr. A* 1000 (2003) 637-655) describe una variedad de escalas de hidrofobia que pueden usarse según la presente invención.

Además de la hidrofobia de aminoácidos, los procedimientos descritos en el presente documento pueden asignar una hidrofobia a un átomo dentro de una proteína o modelo estructural de proteínas. En una realización, la "hidrofobia del átomo" es una relación de la hidrofobia del aminoácido que comprende el átomo y el número de átomos en el aminoácido, o más preferentemente, el número de átomos en la cadena lateral del aminoácido. En una realización similar, la "hidrofobia del átomo" puede ser una fracción de la hidrofobia del residuo que es proporcional al tamaño, área superficial o volumen del átomo en cuestión. Por ejemplo, si un átomo de oxígeno compone el 5% del volumen de un residuo de aminoácido, la hidrofobia del átomo de oxígeno será del 5% de la hidrofobia del residuo de aminoácido. En otra realización, la hidrofobia del átomo puede ser una fracción de la hidrofobia del residuo equivalente a o proporcional a la fracción del área superficial con la que el átomo contribuye al residuo de aminoácido. En realizaciones relacionadas, el peso de la hidrofobia (es decir, la fracción de hidrofobia del residuo) asignado a un átomo puede reflejar la fracción de volumen que requiere el átomo en el residuo, el peso másico del átomo en el residuo, la contribución del átomo a la hidrofobia, etc. Como se ha descrito anteriormente, la hidrofobia de aminoácidos se determina según una escala de hidrofobia conocida en la técnica.

El término "región propensa a la agregación", como se trata en el presente documento, es una región sobre una estructura de proteína que tiene una propensión a unirse a otras proteínas, aumentando así la probabilidad de la formación de agregados. Las regiones propensas a la agregación presentan carácter hidrófobo como se ha identificado por las puntuaciones de SAP descritas en el presente documento. En otra realización, una región propensa a la agregación es una región que es más hidrófoba que las regiones de alrededor. En una realización específica, la región propensa a la agregación puede ser una región espacial definida tridimensional, por ejemplo, una esfera de radio R (o, alternativamente, todos los residuos de aminoácidos con al menos un átomo dentro del radio R), rodeando un átomo en el que el carácter hidrófobo es la puntuación de SAP. En otras realizaciones, la "región propensa a la agregación" engloba cualquier grupo o agrupación de residuos o átomos que presentan un carácter hidrófobo como se calcula por la puntuación de SAP. Alternativamente, una "región propensa a la agregación" puede comprender átomos o residuos próximos que tienen una puntuación de SAP superior a algún umbral, por ejemplo, $>-0,5$, >0 , $>0,5$, etc., o, en una realización similar, puede comprender aquellos átomos o residuos que tienen un área bajo la curva calculada (en una representación de puntuaciones de SAP como se describe más adelante) por encima de algún umbral, por ejemplo, $>-0,5$, >0 , $>0,5$, >1 , $>1,5$, >2 , $>2,5$, etc.

En un aspecto, los procedimientos de la invención emplean tecnología de simulación molecular para preprocesar modelos estructurales de proteínas y/o para identificar regiones propensas a la agregación en proteínas. Por ejemplo, puede emplearse una simulación de dinámica molecular antes de calcular SAP o SAA. En la práctica, cualquier técnica/paquete de simulación que muestree espacio conformacional puede usarse según los procedimientos descritos en el presente documento. El modo preferido de simulación molecular es una simulación de dinámica molecular (MDS). Una MDS es una simulación matemática en la se deja que los átomos en una estructura molecular se muevan e interaccionen según las leyes de la física, por ejemplo, puede permitirse que los enlaces químicos dentro de proteínas se flexionen, roten, doblen o vibren como se permite por las leyes de la química y la física. Interacciones tales como fuerzas electrostáticas, fuerzas hidrófobas, interacciones de van der Waals, interacciones con disolvente y otras también pueden modelarse en simulaciones de MDS. Tales simulaciones permiten que un experto en la materia observe la estructura de la proteína como podría aparecer cuando se solvata, o tomara mediciones más exactas de la estructura de proteína promediando múltiples mediciones en diversos puntos durante la simulación. En una realización preferida, la simulación molecular se realiza usando el paquete de simulación CHARMM (Brooks y col. *J. Comput. Chem.*, 1983,4, 187). En otra realización preferida, la simulación molecular se realiza usando el paquete NAMD (Phillips y col. *Journal of Computational Chemistry*. 2005, 26, 1781). Un experto en la materia entenderá que pueden usarse múltiples paquetes, por ejemplo, puede emplearse el paquete CHARMM para establecer o preprocesar un modelo estructural de proteínas, solvatar la estructura, etc., y puede emplearse el paquete NAMD para las simulaciones que se convierten en parte de los cálculos de la propensión a la agregación espacial. Cualquiera de las numerosas metodologías conocidas en la técnica para realizar simulaciones de MDS puede usarse según la presente invención. Las siguientes publicaciones describen múltiples metodologías que pueden emplearse: Guvench y MacKerell. *Methods Mol Biol*. 2008; 443:63-88; Norberg y Nilsson. *Q Rev Biophys*. Agosto de 2003; 36(3):257-306; las patentes de EE.UU. nº 5424963; 7096167 y las solicitudes de patente de EE.UU. nº 11/520.588; y 10/723.594. En particular, las siguientes plataformas de software pueden emplearse para simulaciones de dinámica molecular: ABINIT (Gonze y col. *Comput. Mat. Science*. 2002, 25, 478; Gonze y col. *Kristallogr*. 2005, 220, 558; abinit.org/); AMBER (Duan y col. *Journal of Computational Chemistry*. 2003, 24(16):1999-2012; amber.scripps.edu); Ascalaph (agilemolecule.com/Products.html, 19 de junio de 2008); CASTEP (Segall y col. *J. Phys.: Cond. Matt*. 2002, 14(11):2717-2743; Clark y col. *Zeitschrift für Kristallographie*. 2005, 220(5-6) pág. 567-570; CASTEP.org); CPMD (manual de CMPD para CMPD versión 3.11.0, 29 de marzo de 2006; cpmd.org/manual.pdf); CHARMM (Brooks y col. *J Comp Chem*. 1983, 4:187-217; charmm.org); DL_POLY (Todorov & Smith, THE DL POLY 3 USER MANUAL. STFC Daresbury Laboratory. Versión 3.09.3, febrero de 2008; cse.scitech.ac.uk/ccg/software/DL_POLY/MANUALS/USRMAN3.09.pdf); FIREBALL

(fireball.phys.wvu.edu/LewisGroup/fireballHome.html); GROMACS (Van Der Spoel y col., J Comput Chem. 2005, 26(16): 1701-18. Hess y col., J Chem Theory Comput. 2008, 4(2): 435; gromacs.org); GROMOS (Schuler, Daura, van Gunsteren. Journal of Computational Chemistry. 2001, 22(11):1205-1218; igc.ethz.ch/GROMOS/index); LAMMPS (Plimpton, J Comp Phys. 1995, 117, 1-19; lammmps.sandia.gov); MDynaMix (Lyubartsev y Laaksonen. Computer Physics Communications. 2000, 128, 565-589; fos.su.se/~sasha/mdynamix/); MOLLY (Moldy: a portable molecular dynamics simulation program for serial and parallel computers, Computer Physics Communications. 2000, 126(3):309-328; earth.ox.ac.uk/~keithr/moldy.html); MOSCITO (Dietmar Paschek y Alfons Geiger. User's Guide and Manual, MOSCITO 4, Performing Molecular Dynamics Simulations, 7 de abril de 2003, ganter.chemie.uni-dortmund.de/MOSCITO/manual4.pdf); NAMD (Kumar y col. IBM Journal of Research and Development. 2007, volumen 52, nº 1/2; Phillips y col., Proceedings of SC 2002; charm.cs.uiuc.edu/research/moldyn/); Newton-X (M. Barbatti, G. Granucci, M. Ruckebauer, M. Persico, H. Lischka, Newton-X: a package for Newtonian dynamics close to the crossing seam, versión 0.15b, 2007; univie.ac.at/newtonx; Barbatti y col., J. Photochem. Photobio. A 190, 228 (2007)); ProtoMol (Matthey y col. ACM Trans. Math. Softw., 2004, 30(3):237-265; protomol.sourceforge.net/); PWscf (guía de usuario para Quantum-ESPRESSO versión 3.2, pwscf.org/guide/3.2.3/users-guide-3.2.3.pdf); SIESTA (Soler y col. Journal of Physics: Condensed Matter. 2002, 14:2745-2779; uam.es/departamentos/ciencias/fismateriac/siesta/); VASP (Georg Kresse y Jürgen Furthmüller, VASP the GUIDE, Institut für Materialphysik, Universidad de Viena, Sensengasse 8, A-1130 Austria, Viena, 1 de marzo de 2007; cms.mpi.univie.ac.at/vasp/); TINKER (Ren y Ponder. J. Phys. Chem. B. 2003, 107, 5933-5947; dasher.wustl.edu/tinker/); YASARA (Krieger E, Koraimann G, Vriend G. Proteins. 2002 47(3):393-402); ORAC (Procacci y col., Fis. Chem. 1996, 100 10464-10469; chim.unifi.it/oac/); XMD (manual online de XMD, XMD - Molecular Dynamics Program Jon Rifkin, v2.5.30 de 20 de enero de 2002)

Como se usa en el presente documento, los términos “aminoácido” y “residuo de aminoácido” y “residuo” pueden, en algunas realizaciones, usarse sinónimamente para referirse a un aminoácido como existe en un estado aislado, por ejemplo, en disolución tienen grupos terminales amino y carboxi sin unir, o como existe en una proteína, por ejemplo, un residuo de aminoácido ligado covalentemente a al menos otro aminoácido mediante un enlace peptídico. Un experto en la materia entenderá la química de proteínas prevista.

Como se usa en el presente documento, un “aminoácido no natural” es un aminoácido que no se sabe que se produzca en la naturaleza. El término “aminoácido no natural” engloba análogos de aminoácidos. Pueden englobar adicionalmente un derivado de un aminoácido natural que comprende una sustitución o adición seleccionada del grupo que comprende un grupo alquilo, un grupo arilo, un grupo acilo, un grupo azido, un grupo ciano, un grupo halo, un grupo hidracina, un grupo hidrazida, un grupo hidroxilo, un grupo alquenoilo, un grupo alquinoilo, un grupo éter, un grupo tiol, un grupo sulfonilo, un grupo seleno, un grupo éster, un grupo tioácido, un grupo borato, un grupo boronato, un grupo fosfo, un grupo fosfono, un grupo fosfina, un grupo heterocíclico, un grupo enona, un grupo imina, un grupo aldehído, un grupo hidroxilamino, un grupo ceto, un grupo de azúcar, un grupo .alfa.-hidroxi, un grupo ciclopropilo, un grupo ciclobutilo, un grupo ciclopentilo, un grupo 2-nitrobencilo, un grupo 3,5-dimetoxi-2-nitrobencilo, un grupo carbamato de 3,5-dimetoxi-2-nitroveratrol, un grupo nitrobencilo, un grupo 3,5-dimetoxi-2-nitrobencilo y un grupo amino.

Por ejemplo, el aminoácido no natural puede ser, sin limitación, cualquiera de los siguientes aminoácidos: hidroximetionina, norvalina, O-metilserina, crotiglicina, hidroxileucina, alo-isoleucina, norleucina, ácido α -aminobutírico, t-butilalanina, hidroxiglicina, hidroxiserina, F-alanina, hidroxitirosina, 2-F-tirosina, 3-F-tirosina, 4-metilfenilalanina, 4-metoxi-fenilalanina, 3-hidroxi-fenilalanina, 4-NH₂-fenilalanina, 3-metoxi-fenilalanina, 2-F-fenilalanina, 3-F-fenilalanina, 4-F-fenilalanina, 2-Br-fenilalanina, 3-Br-fenilalanina, 4-Br-fenilalanina, 2-Cl-fenilalanina, 3-Cl-fenilalanina, 4-C1-fenilalanina, 4-CN-fenilalanina, 2,3-F₂-fenilalanina, 2,4-F₂-fenilalanina, 2,5-F₂-fenilalanina, 2,6-F₂-fenilalanina, 3,4-F₂-fenilalanina, 3,5-F₂-fenilalanina, 2,3-Br₂-fenilalanina, 2,4-Br₂-fenilalanina, 2,5-Br₂-fenilalanina, 2,6-Br₂-fenilalanina, 3,4-Br₂-fenilalanina, 3,5-Br₂-fenilalanina, 2,3-Cl₂-fenilalanina, 2,4-Cl₂-fenilalanina, 2,5-Cl₂-fenilalanina, 2,6-Cl₂-fenilalanina, 3,4-Cl₂-fenilalanina, 2,3,4-F₃-fenilalanina, 2,3,5-F₃-fenilalanina, 2,3,6-F₃-fenilalanina, 2,4,6-F₃-fenilalanina, 3,4,5-F₃-fenilalanina, 2,3,4-Br₃-fenilalanina, 2,3,5-Br₃-fenilalanina, 2,3,6-Br₃-fenilalanina, 2,4,6-Br₃-fenilalanina, 3,4,5-Br₃-fenilalanina, 2,3,4-Cl₃-fenilalanina, 2,3,5-Cl₃-fenilalanina, 2,3,6-Cl₃-fenilalanina, 2,4,6-Cl₃-fenilalanina, 3,4,5-Cl₃-fenilalanina, 2,3,4,5-F₄-fenilalanina, 2,3,4,5-Br₄-fenilalanina, 2,3,4,5-Cl₄-fenilalanina, 2,3,4,5,6-F₅-fenilalanina, 2,3,4,5,6-Br₅-fenilalanina, 2,3,4,5,6-Cl₅-fenilalanina, ciclohexilalanina, hexahidrotirosina, ciclohexanol-alanina, hidroxilalanina, hidroxifenilalanina, hidroxivalina, hidroxisoleucina, hidroxilglutamina, tienilalanina, pirrolalanina, N_T-metil-histidina, ácido 2-amino-5-oxohexanoico, norvalina, norleucina, 3,5-F₂-fenilalanina, ciclohexilalanina, 4-Cl-fenilalanina, p-azido-fenilalanina, o-azido-fenilalanina, O-4-alil-L-tirosina, ácido 2-amino-4-pentanoico y ácido 2-amino-5-oxohexanoico. Se espera que, al menos para los aminoácidos no naturales enumerados anteriormente y para aquellos empleados por la tecnología Ambrx ReCODE™ (ambrx.com/wt/page/technology), los aminoácidos no naturales sigan escalas de hidrofobia similares a las de los 20 aminoácidos comunes, por ejemplo, como se describe en Black y Mould. Alternativamente, la hidrofobia de cualquier aminoácido no natural o inusual puede determinarse por diversas técnicas que son muy conocidas en la técnica, tal como aquellas revisadas y referenciadas en Biswas y col. (*J. Chromatogr. A* 1000 (2003) 637-655).

El término “análogo de aminoácido” se refiere a un aminoácido en el que el grupo carboxi del extremo C, el grupo amino del extremo N o grupo funcional de la cadena lateral se ha modificado químicamente a otro grupo funcional. Por ejemplo, (éster beta-metilico de) ácido aspártico es un análogo de aminoácido de ácido aspártico; N-etilglicina es

un análogo de aminoácido de glicina; o alaninacarboxamida es un análogo de aminoácido de alanina.

El término “aminoácido inusual” se refiere a aquellos aminoácidos naturales que son raros o de otro modo no se encuentran entre los aminoácidos más comunes, siendo los aminoácidos más comunes selenocisteína, alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutámico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina y valina.

Otros ejemplos no limitantes de los aminoácidos modificados, inusuales (es decir, raros), no naturales o análogos que pueden estar sustituidos en una proteína según los procedimientos de la invención son: O-metil-L-tirosina, L-3-(2-naftil)-alanina, 3-metil-L-fenilalanina, fenilalanina fluorada, p-benzoil-L-fenilalanina, p-yodo-L-fenilalanina, p-bromo-L-fenilalanina, p-amino-L-fenilalanina, 3,4-dihidroxi-L-fenilalanina, isopropil-L-fenilalanina, p-azido-L-fenilalanina, p-acetil-L-fenilalanina, m-acetil-L-fenilalanina, 4-(2-oxo-propoxi)-L-fenilalanina y los aminoácidos (y procedimientos de incorporación de los mismos) que se describen en las patentes de EE.UU. nº 7.083.970; 7.045.337; solicitud de patente de EE.UU. nº 10/126.931; 11/002.387; 11/254.170; 11/009.635; 11/670.354; 11/284.259; 10/563.686; 11/326.970; 10/563.656; 10/563.655; 11/715.672; 11/671.036; 11/255.601; 11/580.223; 11/137.850; 11/233.508; 10/575.991; 11/232.425; publicaciones WIPO WO/2007/094916; WO/2007/130453; y las publicaciones Liao J. Biotechnol Prog. Enero-Febrero de 2007; 23(1):28-31; Rajesh, y Iqbal. Curr Pharm Biotechnol. Agosto de 2006; 7(4):247-59. Cardillo y col. Mini Rev Med Chem. Marzo de 2006; 6(3):293-304; Wang y col. Annu Rev Biophys Biomol Struct. 2006; 35:225-49; Chakraborty y col., Glycoconj J. Marzo de 2005; 22(3):83-93. Otros ejemplos de aminoácidos no naturales pueden encontrarse, por ejemplo, en las siguientes publicaciones de patente de EE.UU.: 2003-0082575, 2005-0250183, 2003-0108885, 2005-0208536 y 2005-0009049.

I. Propensión a la agregación espacial

La invención en el presente documento se refiere a procedimientos para identificar regiones propensas a la agregación sobre una superficie de proteína, para prevenir o reducir la agregación de una proteína, y para identificar una región de unión a macromolécula sobre una proteína. Los procedimientos en el presente documento representan un avance en la capacidad de los procedimientos computacionales para identificar regiones de proteína que pueden modificarse para reducir la propensión de una proteína a agregarse o para reducir la afinidad de unión de una proteína por una macromolécula. En particular, los procedimientos se basan, al menos en parte, en el cálculo del SAA (área accesible al disolvente), que se conoce en la técnica por caracterizar la superficie de una proteína. El SAA da el área superficial de cada aminoácido o estructura de proteína que está en contacto con el disolvente. El SAA puede calcularse normalmente calculando el sitio del centro de una esfera de sonda a medida que rueda sobre la superficie de la proteína, es decir, la superficie de un modelo estructural de proteínas. La esfera de sonda tiene el mismo radio que el de una molécula de agua, $R=1,4 \text{ \AA}$. Procedimientos alternativos de cálculo de SAA, descritos más adelante, se conocen en la técnica y son compatibles con los procedimientos descritos en el presente documento. Aunque el SAA es bastante útil para caracterizar la superficie de la proteína, no se encontró que fuera adecuado para caracterizar los parches hidrófobos sobre la superficie de proteína que son posiblemente propensos a la agregación debido a los siguientes fallos,

1. SAA no distingue entre regiones hidrófobas e hidrófilas
2. SAA no es directamente proporcional a una hidrofobia del residuo (por ejemplo, MET tiene más área superficial que LEU pero es menos hidrófoba)
3. SAA no indica si varios residuos hidrófobos están cerca y así podría potenciar la hidrofobia de una cierta región. Estos residuos podrían estar cerca tanto en la secuencia primaria como en la estructura terciaria aún cuando estuvieran lejos en la secuencia primaria. De cualquier forma, podrían potenciar la hidrofobia de un cierto parche sobre la superficie del anticuerpo.

Una medida que se describe en el presente documento, el SAA eficaz, se genera calculando la hidrofobia de la fracción del aminoácido que se expone según la siguiente fórmula:

$$\text{SAA eficaz} - \text{SAA} = \frac{\text{SAA}}{\text{SAA}_{\text{completamente expuesta}}} \times \text{Hidrofobia del residuo}$$

Otra realización del SAA eficaz comprende además sumar el SAA eficaz durante al menos dos, al menos tres, al menos cuatro, al menos cinco o al menos seis (por ejemplo, dos, tres, cuatro, cinco, seis, etc.) residuos de aminoácidos que son adyacentes en la secuencia de proteínas primaria. Aunque el SAA eficaz representa una mejora con respecto al SAA básica, sin embargo carece de la capacidad para explicar completamente la estructura de la proteína plegada y el hecho de que aminoácidos que no son adyacentes en la secuencia de proteínas puedan estar en proximidad entre sí en la estructura secundaria, terciaria o cuaternaria de una proteína. Tales pliegues de proteína pueden formar regiones propensas a la agregación que no aparecen en la estructura primaria sola, o que solo pueden detectarse analizando más robustamente la estructura de la proteína plegada.

La presente invención proporciona una nueva medida más avanzada, llamada la propensión a la agregación espacial, que resaltará la eficaz hidrofobia de un cierto parche o región sobre la superficie de la proteína. La

propensión a la agregación espacial se calcula para regiones espaciales definidas sobre o próximas a los átomos de un modelo estructural de proteínas.

En este contexto, una "región espacial definida" es un espacio o volumen tridimensional elegido para capturar una estructura física local y/o entorno químico sobre o próximo a la estructura de la proteína. En una realización particularmente preferida, la propensión a la agregación espacial se calcula para regiones esféricas con radio R centrado en átomos en una proteína (por ejemplo, átomos en un modelo estructural de proteínas). La propensión a la agregación espacial también puede calcularse para regiones esféricas con radio R centrado en enlaces químicos, o posicionado en el espacio próximo al modelo estructural. Por consiguiente, en otra realización preferida, la SAP puede calcularse para una región espacial definida centrada próxima a un átomo, por ejemplo, centrada en un punto en el espacio que está entre 1-10 Å, más preferentemente 1-5 Å, más preferentemente 1-2 Å, del centro de un átomo particular o enlace químico.

En realizaciones preferidas, el radio R elegido está entre 1 Å y 50 Å, más preferentemente entre 1 Å y 50 Å. En realizaciones particulares, el radio elegido es al menos 1 Å, al menos 3 Å, al menos 4 Å, al menos 5 Å, al menos 6 Å, al menos 7 Å, al menos 8 Å, al menos 9 Å, al menos 10 Å, al menos 11 Å, al menos 12 Å, al menos 15 Å, al menos 20 Å, al menos 25 Å, o al menos 30 Å. En realizaciones particularmente preferidas, el radio elegido está entre 5 Å y 15 Å, más preferentemente entre 5 Å y 12 Å, más preferentemente entre 5 Å y 10 Å. En realizaciones específicas, el radio elegido es 5 Å o 10 Å.

En otras realizaciones, la región para que la propensión a la agregación espacial se calcula no es esférica. La posible forma de la región puede comprender además un cubo, un cilindro, un cono, esferoide elíptico, una pirámide, una semiesfera, o cualquier otra forma que pueda usarse para encerrar una parte del espacio. En tales realizaciones, el tamaño de la región puede elegirse usando medidas distantes del radio, por ejemplo, la distancia del centro de la forma a una cara o vértice.

En una realización preferida, la SAP puede usarse para seleccionar residuos en una proteína que pueden estar sustituidos, aumentando así la estabilidad de la proteína. En estudios previos, dos enfoques principales para estabilizar una proteína *in vitro* han sido (1) manipular la propia secuencia de proteínas y (2) incluir aditivos en la formulación líquida. Ambos enfoques se han investigado y se han obtenido resultados significativos. El primer enfoque se ha basado en cribar amplias bibliotecas de variantes aleatorias por ordenador o experimentalmente. En el segundo enfoque, cribado de alta resolución para estabilizar aditivos, además del diseño racional de aditivos, permite la identificación de formulaciones óptimas para una proteína terapéutica.

Se espera que la presente invención modernice el procedimiento de potenciamiento de la estabilidad identificando puntos calientes existentes para agregación computacionalmente y analizando variantes con sustituciones en aquellos sitios experimentalmente.

Así, en términos generales, un procedimiento para calcular la propensión a la agregación espacial para un átomo particular en una proteína comprende (a) identificar uno o más átomos en un modelo estructural que representa la proteína, en el que el uno o más átomos están dentro de una región espacial definida centrada en o próxima al átomo particular; (b) calcular, para cada uno del uno o más átomos en la región espacial definida, una relación del área accesible al disolvente (SAA) de los átomos con respecto al SAA de átomos en un residuo idéntico que está completamente expuesto; (c) multiplicar cada relación por la hidrofobia del átomo del uno o más átomos; y (d) sumar los productos de la etapa (c); por lo que la suma es la SAP para el átomo particular.

En una realización relacionada, la SAP puede calcularse según un procedimiento diferente que comprende (a) identificar uno o más residuos de aminoácidos en un modelo estructural que representa la proteína, en el que el uno o más residuos de aminoácidos tienen al menos un átomo dentro de una región espacial definida centrada en o próxima al átomo particular; (b) calcular, para cada uno de los uno o más residuos de aminoácidos identificados, una relación del área accesible al disolvente (SAA) de átomos en el aminoácido con respecto al SAA de átomos en un residuo idéntico que está completamente expuesto; (c) multiplicar cada relación por la hidrofobia del uno o más residuos de aminoácidos como se ha determinado por una escala de hidrofobia de aminoácidos; y (d) sumar los productos de la etapa (c); por lo que la suma es la SAP para el átomo particular. En realizaciones preferidas, el modelo estructural se procesa antes de la etapa (a) permitiendo que el modelo estructural interactúe con disolvente en una simulación de dinámica molecular. Si se identifica que un aminoácido tiene al menos un átomo dentro de la región espacial definida, puede requerirse que el al menos un átomo sea exclusivamente un átomo en un cadena lateral del aminoácido. Alternativamente puede ser un átomo que se requiere que sea un átomo de la cadena principal.

En otras realizaciones, este procedimiento puede comprender además opcionalmente realizar una simulación de dinámica molecular antes de la etapa (a) y repetir las etapas (a)-(d), realizando cada vez otra simulación de dinámica molecular en una pluralidad de etapas de tiempo, produciendo así múltiples sumas como en la etapa (d), y calcular el promedio de las sumas; por lo que el promedio calculado es la SAP para el átomo particular.

En otras realizaciones preferidas, la SAP puede usarse para seleccionar residuos en una proteína que puede estar

sustituida, reduciendo así la afinidad de unión de la proteína por una macromolécula.

Un experto en la materia apreciará que una realización de la presente invención que emplea el promedio de valores calculados durante una simulación de dinámica molecular será computacionalmente más intensa. Por tanto, en algunos casos, una realización tal proporcionará un mapa más preciso o altamente resuelto de la propensión a la agregación espacial. Sin embargo, experimentos tratados en el presente documento han mostrado que el procedimiento todavía es altamente exacto cuando no se emplea el promedio de dinámica molecular. En una realización preferida, los valores de propensión a la agregación espacial pueden calcularse para todas las estructuras de proteína en una base de datos, por ejemplo, la base de datos de proteínas (PDB), identificando así rápidamente residuos y parches hidrófobos sobre todas las estructuras de proteína conocidas. Este procedimiento permite el rápido cribado de grandes conjuntos de proteínas para identificar posibles regiones propensas a la agregación y/o sitios de interacción de proteínas.

En una aplicación preferida, la propensión a la agregación espacial se describe por la siguiente fórmula:

$$\left(\text{Pr opensión a la agregación} \right)_{\text{espacial (SAP)}}^{\text{átomo } i} = \sum_{\text{Simulación Promedio}} \left\{ \sum_{\text{átomos dentro de R del átomo } i} \left(\frac{\text{SAA de átomos de la cadena lateral dentro del radio R}}{\text{SAA de átomos de la cadena lateral del residuo completamente expuesto}} \times \text{Hidrofobia del átomo} \right) \right\}$$

en la que en

- 1) SAA de átomos de la cadena lateral dentro del radio R se calcula en cada instante de la simulación. SAA se calcula preferentemente en el modelo de simulación calculando el sitio del centro de una esfera de sonda cuando rueda sobre la superficie de la proteína. La esfera de sonda tiene el mismo radio que una molécula de agua, R=1,4 Å. Un experto en la materia apreciará que otros procedimientos de cálculo de SAA serían compatibles con los procedimientos descritos aquí para calcular SAP. Por ejemplo, el SAA puede calcularse sobre solo los átomos de la cadena lateral del aminoácido. El SAA también puede calcularse sobre solo los átomos de la cadena principal del aminoácido (es decir, aquellos átomos del esqueleto de péptido e hidrógenos asociados). Alternativamente, el SAA puede calcularse sobre solo los átomos de la cadena principal de aminoácidos con la exclusión de hidrógenos asociados;
- 2) SAA de la cadena lateral del residuo completamente expuesto (dicho para el aminoácido 'X') se obtiene, en una realización preferida, calculando el SAA de cadenas laterales del residuo central en la conformación completamente extendida del tripéptido 'Ala-X-Ala'; y
- 3) Hidrofobia del átomo se obtiene como se ha descrito anteriormente usando la escala de hidrofobia de Black y Mould (Black y Mould, Anal. Biochem. 1991, 193, 72-82).

Un residuo que está "completamente expuesto" es un residuo, X, en la conformación completamente extendida del tripéptido Ala-X-Ala. Un experto en la materia apreciará que esta disposición se diseña de forma que un cálculo de SAA sobre un residuo tal, X, dará la máxima área accesible al disolvente disponible. Por consiguiente, se contempla que otros residuos, además de la alanina, puedan usarse en el cálculo sin afectar o alterar completamente los resultados.

Como se ha descrito anteriormente, los procedimientos de la presente invención pueden aplicarse a cualquier modelo estructural de proteínas. Por consiguiente, la SAP basada en solo la estructura de rayos X puede exponerse como:

$$\left(\text{Pr opensión a la agregación} \right)_{\text{espacial (SAP)}}^{\text{rayos X}} = \sum_{\text{Simulación Promedio}} \left\{ \sum_{\text{átomos dentro de R del átomo } i} \left(\frac{\text{SAA de átomos de la cadena lateral dentro del radio R}}{\text{SAA de átomos de la cadena lateral del residuo completamente expuesto}} \times \text{Hidrofobia del átomo} \right) \right\}$$

Similarmente, si la estructura de rayos X no está disponible, el mismo parámetro de propensión a la agregación espacial puede aplicarse a la estructura generada mediante modelado por homología, y el parámetro de SAP puede así exponerse como:

$$\left(\text{Pr opensión a la agregación} \right)_{\text{espacial (SAP)}}^{\text{estructura por homología}} = \sum_{\text{Simulación Promedio}} \left\{ \sum_{\text{átomos dentro de R del átomo } i} \left(\frac{\text{SAA de átomos de la cadena lateral dentro del radio R}}{\text{SAA de átomos de la cadena lateral del residuo completamente expuesto}} \times \text{Hidrofobia del átomo} \right) \right\}$$

En realizaciones preferidas, la propensión a la agregación espacial se calcula para todos los átomos en un modelo estructural de proteínas. En algunas realizaciones, los valores atómicos de la propensión a la agregación espacial pueden promediarse sobre cada residuo de proteína individual, o sobre pequeños grupos de residuos.

II. Usos de la invención

En un aspecto, la presente invención puede usarse como se ha descrito anteriormente para identificar residuos de aminoácidos hidrófobos, regiones o parches en una proteína. Sin querer detenerse en valores umbral específicos, átomos o residuos de aminoácidos que tienen una propensión a la agregación espacial > 0 se consideran que son hidrófobos, o que están en una región propensa a la agregación. Dependiendo del tipo de proteína, la estructura particular y el disolvente en el que existe, puede desearse identificar átomos o residuos usando un corte que es ligeramente inferior a cero, por ejemplo, eligiendo átomos o residuos que tienen una propensión a la agregación espacial superior a $-0,1$, $-0,15$, $-0,2$, etc. Alternativamente, puede desearse emplear un corte más riguroso, por ejemplo, 0 , $0,05$, $0,1$, $0,15$, $0,2$, etc., con el fin de elegir los átomos hidrófobos más fuertes, residuos o parches. En otra realización, puede ser ventajoso simplemente seleccionar átomos o residuos que tienen propensión a la agregación espacial que son mayores que átomos o residuos que son próximos tanto secuencialmente (es decir, a lo largo de la secuencia de proteínas) como, en una realización preferida, espacialmente (es decir, en la estructura tridimensional). Un procedimiento preferido para seleccionar átomos o residuos en un parche hidrófobo es mapear los valores de propensión a la agregación espacial calculados, por ejemplo, usando una codificación de colores o codificación numérica, sobre el modelo estructural de proteínas del que se derivaron, visualizando así diferencias en la propensión a la agregación espacial a través de la superficie de proteína y de ahí permitiendo la fácil selección de parches o residuos hidrófobos. En una realización particularmente preferida, los cálculos para propensión a la agregación espacial se llevan a cabo por separado usando dos valores elegidos para el radio, uno de mayor resolución, por ejemplo, 5 \AA , y uno de menor resolución, por ejemplo, 10 \AA . En una realización tal pueden observarse parches hidrófobos mayores o más anchos sobre la estructura de proteína con el menor mapa de resolución. Una vez se han seleccionado parches hidrófobos de interés sobre el mapa de baja resolución, aquellos parches pueden visualizarse en mayor detalle en el mapa de mayor resolución que puede, en algunas realizaciones, permitir que un experto en la materia elija más fácilmente o con más exactitud residuos para mutar o modificar. Por ejemplo, si se visualiza un parche hidrófobo en el mapa de mayor resolución, puede desearse seleccionar para la mutación el residuo que tiene la mayor puntuación de SAP o que es el más hidrófobo (por ejemplo, el residuo más hidrófobo en el parche según la escala de Black y Mould, Anal. Biochem. 1991, 193, 72-82).

En una realización específica, un procedimiento para identificar una región propensa a la agregación sobre una proteína comprende (a) mapear, sobre el modelo estructural, la SAP para átomos en la proteína como se calcula para un átomo particular según cualquiera de los procedimientos descritos en el presente documento; y (b) identificar una región dentro de la proteína que tiene una pluralidad de átomos que tiene una SAP > 0 ; en el que la región propensa a la agregación comprende los aminoácidos que comprenden dicha pluralidad de átomos. En una realización tal, la SAP puede calcularse para todos los átomos en una proteína o una parte de los átomos. Se contempla que solo puede calcularse la SAP para residuos particulares o grupos de residuos que son de interés.

En una realización similar, puede ser informativo representar las puntuaciones de SAP de los átomos (o la puntuación de SAP como se promedia sobre residuos de aminoácidos). Una representación tal que muestra la puntuación de SAP a lo largo de los átomos o residuos de una proteína permite la fácil identificación de picos, que pueden indicar candidatos para sustitución. En una realización particularmente preferida, las puntuaciones de SAP a lo largo de los átomos o residuos en la proteína se representan en una gráfica y el área bajo la curva (ABC) se calcula para picos en la gráfica. En una realización tal, picos con un ABC mayor representan regiones propensas a la agregación grandes o más hidrófobas. En realizaciones particulares se deseará seleccionar para la sustitución uno o más residuos que se identifican como existentes en un pico, o, más preferentemente, en un pico con un ABC grande.

En realizaciones particulares, la presente invención puede usarse para preparar una variante de proteína que presenta una propensión reducida a la agregación sustituyendo al menos un residuo de aminoácido dentro de una región propensa a la agregación en la proteína identificada por cualquiera de los procedimientos descritos en el presente documento con un residuo de aminoácido que es más hidrófilo que el residuo que está siendo sustituido, de forma que se reduzca la propensión a la agregación de la variante. Como se usa en el presente documento, cuando los residuos de aminoácidos se denominan "más" o "menos" hidrófilos o hidrófobos, se apreciará por el experto que esto significa más o menos hidrófobos con respecto a otro aminoácido según una medida de hidrofobia (hidrofilia) conocida en la técnica, por ejemplo, la escala de hidrofobia de Black y Mould.

En una realización similar, la presente invención puede usarse para preparar una variante de proteína que presenta una propensión reducida a la agregación generando una pluralidad de variantes de proteína sustituyendo en cada variante al menos un residuo dentro de una región propensa a la agregación en la proteína, en la que la región propensa a la agregación se identifica usando puntuaciones de SAP calculadas según cualquier procedimiento descrito en el presente documento, en la que un residuo o residuos diferentes, o combinaciones diferentes de residuos, se sustituyen en cada variante, y en la que el al menos un residuo se sustituye con un residuo que es más hidrófilo; y (b) seleccionar una variante de proteína preparada como en (a) que presenta una propensión reducida a la agregación.

Además, un residuo de aminoácido en una región propensa a la agregación puede delecionarse en vez de sustituirse. En algunas proteínas en las que múltiples residuos de aminoácidos se seleccionan para sustitución,

algunos residuos pueden sustituirse mientras que otros se delecionan.

5 En otras realizaciones, múltiples regiones propensas a la agregación o residuos pueden identificarse en una proteína inicial mediante los procedimientos descritos anteriormente (por ejemplo, usando un corte de propensión a la agregación espacial por encima del cual se seleccionan residuos). Posteriormente, una pluralidad de variantes de proteína puede generarse sustituyendo en dicha proteína inicial uno o más residuos de aminoácidos seleccionados (o uno o más residuos que se encuentran en el parche seleccionado) con residuos de aminoácidos que son más hidrófilos, de forma que se cree una pluralidad de variantes de proteína que representa una variedad de diferentes sustituciones de aminoácidos. Esta población puede entonces cribarse para seleccionar una o más variantes de proteína que tienen una propensión reducida a la agregación. Un experto en la materia apreciará que pueden 10 identificarse múltiples regiones propensas a la agregación, y que una o más sustituciones y/o delecciones pueden hacerse en una o más regiones propensas a la agregación. La hidrofobia relativa de los aminoácidos puede determinarse por la escala de hidrofobia de Black y Mould como se ha descrito anteriormente. En realizaciones específicas, un aminoácido que va a sustituirse se selecciona del grupo que comprende o está constituido por Phe, Leu, Ile, Tyr, Trp, Val, Met, Pro, Cys, Ala o Gly. En realizaciones relacionadas, el aminoácido más hidrófilo que se 15 sustituirá en la proteína se elegirá del grupo que comprende o está constituido por Thr, Ser, Lys, Gln, Asn, His, Glu, Asp y Arg.

20 Pueden prepararse variantes de proteína mediante cualquier procedimiento conocido en la técnica que incluye mutagénesis dirigida a sitio y otra tecnología de ADN recombinante, por ejemplo, véanse las patentes de EE.UU. nº 5284760; 5556747; 5789166; 6878531, 5932419; y 6391548.

25 En realizaciones particulares, la presente invención puede usarse para preparar una variante de proteína que presenta una propensión reducida a la agregación sustituyendo al menos un residuo de aminoácido dentro de una región propensa a la agregación en la proteína identificada por cualquiera de los procedimientos descritos en el presente documento con un residuo de aminoácido natural, un residuo de aminoácido modificado, un residuo de aminoácido inusual, un residuo de aminoácido no natural, o un análogo o derivado de aminoácido que es más hidrófilo que el residuo que está siendo sustituido, de forma que se reduzca la propensión a la agregación de la variante.

30 La síntesis de aminoácidos no naturales es conocida para aquellos expertos en la materia, y se describe adicionalmente, por ejemplo, en la publicación de patente de EE.UU. nº 2003-0082575. En general, puede emplearse cualquier procedimiento conocido en la técnica para sintetizar o incorporar aminoácidos no naturales, modificados o inusuales en proteínas que incluyen, pero no se limita a, aquellos procedimientos descritos o 35 referenciados en las publicaciones Liao J. *Biotechnol Prog.* Enero-Febrero de 2007; 23(1):28-31; Rajesh e Iqbal. *Curr Charm Biotechnol.* Agosto de 2006; 7(4):247-59; Cardillo y col. *Mini Rev Med Chem.* Marzo de 2006;6(3):293-304; Wang y col. *Annu Rev Biophys Biomol Struct.* 2006; 35:225-49; Chakraborty y col., y *Glycoconj J.* Marzo de 2005; 22(3):83-93. Como otro ejemplo, la tecnología Ambrx ReCODE™ puede emplearse para desarrollar e incorporar aminoácidos no naturales, o aminoácidos inusuales en proteínas como se indica mediante los 40 procedimientos descritos en el presente documento.

45 Las variantes de proteína según la invención pueden presentar estabilidad potenciada o mejorada como se ha determinado, por ejemplo, por estudios de estabilidad acelerada. Estudios de estabilidad acelerada a modo de ejemplo incluyen, pero no se limitan a, estudios que caracterizan temperaturas de almacenamiento elevadas. Una disminución en la formación de agregados observada para una variante de proteína con respecto a la proteína natural o inicial indica una estabilidad elevada. La estabilidad de variantes de proteína también puede probarse midiendo el cambio en la transición de la temperatura de fusión de una variante con respecto a la proteína natural o inicial. En una realización tal, la estabilidad elevada sería evidente como un aumento en la transición de la temperatura de fusión en la variante. Procedimientos adicionales para medir la agregación de proteínas se describen 50 en la solicitud de patente de EE.UU. nº 10/176.809.

55 En otro aspecto de la invención, la propensión a la agregación espacial calculada puede usarse para identificar sitios de interacción proteína-proteína sobre la superficie de una estructura de proteína. Se conoce en la técnica que los sitios de interacción de proteínas frecuentemente contienen residuos hidrófobos o parches hidrófobos. Se espera que los procedimientos descritos en el presente documento sean útiles en la localización de sitios de unión identificando parches hidrófobos. Tales parches hidrófobos serán entonces candidatos para sitios de reconocimiento de proteína-proteína o proteína-ligando.

60 En otro aspecto, la invención también incluye procedimientos para identificar una región de unión a macromolécula sobre una proteína que comprende (a) mapear, sobre un modelo estructural de la proteína la SAP, átomos en la proteína como se calcula para un átomo particular según uno cualquiera de los aspectos precedentes; y (b) identificar una región dentro de la proteína que tiene una pluralidad de átomos que tiene una SAP > 0; en el que la región de unión a macromolécula comprende los aminoácidos que comprenden dicha pluralidad de átomos.

65 En otro aspecto, la invención incluye procedimientos para identificar una región de unión a macromolécula sobre una proteína que comprenden identificar uno o más aminoácidos que contienen uno o más átomos que tienen una SAP

superior a un umbral elegido; en los que la SAP se calcula según el procedimiento de uno cualquiera de los aspectos previos y en los que la región de unión a macromolécula comprende los aminoácidos identificados.

5 En otro aspecto, la invención incluye procedimientos para identificar una región de unión a macromolécula sobre una proteína que comprenden representar los valores de SAP como se calcula en uno cualquiera de los aspectos precedentes, calcular, para picos en la representación, el área bajo la curva (ABC) e identificar una o más regiones de proteína con un ABC positiva, en los que la región de unión a macromolécula comprende las regiones de proteína identificadas.

10 En otro aspecto, la invención puede usarse para preparar una variante de proteína que presenta una afinidad de unión reducida para una macromolécula que comprende sustituir o delecionar al menos un residuo de aminoácido dentro de una región de unión a macromolécula por la macromolécula en la proteína, en el que la región de unión a macromolécula se identifica usando puntuaciones de SAP calculadas según uno cualquiera de los aspectos previos; y en el que, si el residuo de aminoácido se sustituye, se sustituye con un residuo de aminoácido que es más hidrófilo, de forma que se reduzca la afinidad de unión por la macromolécula de la variante. En ciertas realizaciones, al menos un residuo se sustituye y al menos un residuo se deleciona. En otro aspecto, la invención también incluye procedimientos de preparación de una variante de proteína que presenta una afinidad de unión alterada para una macromolécula que comprenden (a) generar una pluralidad de variantes de proteína sustituyendo en cada variante al menos un residuo dentro de una región de unión a macromolécula para la macromolécula en la proteína, en los que la región de unión a macromolécula se identifica usando puntuaciones de SAP calculadas según uno cualquiera de los aspectos precedentes, en los que un residuo o residuos diferentes, o combinaciones diferentes de residuos, se sustituyen en cada variante; y (b) seleccionar una variante de proteína preparada como en (a) que presenta una afinidad de unión alterada por la macromolécula. En ciertas realizaciones, el al menos un residuo de aminoácido dentro de la región de unión a macromolécula es el residuo más hidrófobo en la región de unión a macromolécula.

20 En ciertas realizaciones, el al menos un residuo de aminoácido dentro de una región propensa a la agregación es Phe, Leu, Ile, Tyr, Trp, Val, Met, Pro, Cys, Ala o Gly. En ciertas realizaciones, el residuo de aminoácido que es más hidrófilo está seleccionado del grupo constituido por Thr, Ser, Lys, Gln, Asn, His, Glu, Asp y Arg. En ciertas realizaciones, el residuo de aminoácido que es más hidrófilo es un aminoácido inusual, no natural o modificado. En ciertas realizaciones, el residuo de aminoácido que es más hidrófilo se determina según la escala de hidrofobia de Black y Mould. En ciertas realizaciones, al menos dos residuos de aminoácidos se sustituyen dentro de la región de unión a macromolécula. En ciertas realizaciones, al menos tres residuos de aminoácidos se sustituyen dentro de la región de unión a macromolécula. En ciertas realizaciones, al menos un residuo se sustituye dentro de más de una región propensa a la agregación dentro de la proteína. En ciertas realizaciones, la región propensa a la agregación se identifica según el procedimiento de uno cualquiera de los aspectos precedentes para identificar una región propensa a la agregación sobre una proteína. En ciertas realizaciones, que pueden combinarse con las realizaciones precedentes, la macromolécula es otra proteína, un polinucleótido o un polisacárido. En ciertas realizaciones, que pueden combinarse con las realizaciones precedentes, la proteína está seleccionada del grupo constituido por un anticuerpo, un fragmento Fab, un fragmento Fab', un fragmento Fd, un fragmento Fv, un fragmento F(ab')₂ y un fragmento Fc. En ciertas realizaciones, que pueden combinarse con las realizaciones precedentes, la proteína es una citocina, una quimiocina, una lipocina, una miocina, un neurotransmisor, una neurotrofina, una interleucina o un interferón. En ciertas realizaciones, que pueden combinarse con las realizaciones precedentes, la proteína es una hormona o factor de crecimiento. En ciertas realizaciones, la macromolécula es un receptor de hormona o receptor de factor de crecimiento. En ciertas realizaciones, la proteína es un receptor o dominio de receptor. En ciertas realizaciones, la macromolécula es un agonista de receptor o un receptor antagonista del receptor o dominio de receptor. En ciertas realizaciones, que pueden combinarse con las realizaciones precedentes, la proteína es un neurotransmisor o neurotrofina. En ciertas realizaciones, la macromolécula es un receptor de neurotransmisor o receptor de neurotrofina.

50 En algunas realizaciones, la divulgación se refiere además a código informático para determinar SAP según los procedimientos de la invención. En otras realizaciones, la invención se refiere a un ordenador, un superordenador o grupo de ordenadores dedicado a realizar los procedimientos de la invención. En todavía otro aspecto, la divulgación describe un servicio basado en web, basado en servidor o basado en internet para determinar regiones propensas a la agregación sobre una proteína, comprendiendo el servicio aceptar los datos sobre una proteína (por ejemplo, un modelo estructural de proteínas) de un usuario (por ejemplo, mediante internet) o recuperar tales datos de una base de datos de forma que el proveedor del servicio pueda generar, recuperar o acceder a una estructura estática de la proteína, que opcionalmente incluye modelado de dinámica molecular de la proteína para proporcionar una estructura dinámica de la proteína, determinar SAP para átomos o residuos de la proteína basados en la estructura estática o dinámica así generada y devolver los datos de SAP, por ejemplo, como modelo estructural mapeado con dichos datos de SAP por el proveedor del servicio, a un usuario. En algunas realizaciones, el usuario es una persona. En otras realizaciones, el usuario es un sistema de ordenador o algoritmo informático automatizado.

65 En algunas realizaciones, la presente divulgación describe un sistema de cálculo de SAP que comprende: un servidor web para proporcionar que un servicio web calcule SAP para una terminal de usuario mediante internet; una base de datos para guardar información general sobre el procedimiento de cálculo, hidrofobia de aminoácidos, etc., y un servidor de cálculo para realizar el cálculo de SAP basándose en información en la base de datos e información proporcionada o transmitida por internet por el usuario.

En algunas realizaciones, el servidor web y el servidor de cálculo son el mismo sistema informático. En algunas realizaciones, el sistema informático es un superordenador, un grupo de ordenadores o una única estación de trabajo o servidor.

5 En una realización relacionada, el servidor web del sistema de cálculo de SAP comprende además un controlador para controlar la operación completa, una unidad de conexión a red para la conexión a internet y una unidad de servidor web para proporcionar un servicio web para calcular SAP para la terminal de usuario conectada mediante internet.

10 Además, realizaciones de la presente invención se refieren adicionalmente a productos de almacenamiento informático con un medio legible por ordenador que contiene código de programa para realizar diversas operaciones implementadas por ordenador, por ejemplo, calcular la SAP para un modelo estructural, calcular SAA, calcular SAA eficaz, manipular modelos estructurales, implementar simulaciones de dinámica molecular, organizar y guardar datos relevantes, o realizar otras operaciones descritas en el presente documento. El medio legible por ordenador es cualquier dispositivo de almacenamiento de datos que pueda guardar los datos que pueden después ser leídos por un sistema informático. Ejemplos de medios legibles por ordenador incluyen, pero no se limitan a, discos duros, discos blandos, memorias flash, discos ópticos (por ejemplo, CD, DVD, HD-DVD, discos Blu-Ray, etc.) y dispositivos de hardware especialmente configurados tales como circuitos integrados específicos para aplicación (ASIC) o dispositivos lógicos programables (PLD). El medio legible por ordenador también puede distribuirse como una señal de datos plasmada en una onda portadora durante una red de sistemas informáticos acoplados de manera que el código legible por ordenador se almacene y ejecute en un modo distribuido. Se apreciará por aquellos expertos en la materia que los elementos de hardware y software anteriormente descritos son de diseño y construcción estándar. Las realizaciones de ordenador, internet, servidor y servicio relacionadas descritas anteriormente pueden aplicarse adicionalmente con respecto al SAA y al SAA eficaz, además de SAP.

25

III. Composiciones farmacéuticas que contienen péptidos y variantes de péptido de la invención

En otro aspecto, la presente invención proporciona una composición, por ejemplo, una composición farmacéutica, que contiene una o más variantes de proteína producidas mediante los procedimientos de la invención, formuladas junto con un vehículo farmacéuticamente aceptable. Las composiciones farmacéuticas de la invención también pueden administrarse en terapia de combinación, es decir, combinadas con otros agentes. Por ejemplo, la terapia de combinación puede incluir una proteína de la presente invención combinada con al menos otro agente anticancerígeno.

35 Como se usa en el presente documento, "vehículo farmacéuticamente aceptable" incluye todos y cada uno de los disolventes, medios de dispersión, recubrimientos, agentes antibacterianos y antifúngicos, agentes isotónicos y de retardo de la absorción y similares que son fisiológicamente compatibles. Preferentemente, el vehículo es adecuado para administración intravenosa, intramuscular, subcutánea, parenteral, espinal o epidérmica (por ejemplo, mediante inyección o infusión). Dependiendo de la vía de administración, el compuesto activo, es decir, la proteína o variante de la misma de la invención, puede recubrirse en un material para proteger el compuesto de la acción de ácidos y otras condiciones naturales que puedan inactivar el compuesto.

Los compuestos farmacéuticos de la invención pueden incluir una o más sales farmacéuticamente aceptables. Una "sal farmacéuticamente aceptable" se refiere a una sal que retiene la actividad biológica deseada del compuesto parental y no confiere ningún efecto toxicológico no deseado (véase, por ejemplo, Berge, S.M. y col. (1977) *J. Pharm. Sci.* 66:1-19). Ejemplos de tales sales incluyen sales de adición de ácido y sales de adición de base. Las sales de adición de ácido incluyen aquellas derivadas de ácidos inorgánicos no tóxicos tales como clorhídrico, nítrico, fosfórico, sulfúrico, bromhídrico, yodhídrico, fosforoso y similares, además de ácidos orgánicos no tóxicos tales como ácidos mono- y dicarboxílicos alifáticos, ácidos alcanicos sustituidos con fenilo, ácidos hidroxialcanoicos, ácidos aromáticos, ácidos sulfónicos alifáticos y aromáticos y similares. Sales de adición de base incluyen aquellas derivadas de metales alcalinotérreos tales como sodio, potasio, magnesio, calcio y similares, además de aminas orgánicas no tóxicas tales como N,N'-dibenciletilendiamina, N-metilglucamina, cloroprocaína, colina, dietanolamina, etilendiamina, procaína y similares.

55 Una composición farmacéutica de la invención también puede incluir un antioxidante farmacéuticamente aceptable. Ejemplos de antioxidantes farmacéuticamente aceptables incluyen: (1) antioxidantes solubles en agua tales como ácido ascórbico, clorhidrato de cisteína, bisulfato de sodio, metabisulfito de sodio, sulfito de sodio y similares; (2) antioxidantes solubles en aceite tales como palmitato de ascorbilo, hidroxianisol butilado (BHA), hidroxitolueno butilado (BHT), lecitina, galato de propilo, alfa-tocoferol y similares; y (3) agentes quelantes metálicos tales como ácido cítrico, ácido etilendiaminatetraacético (EDTA), sorbitol, ácido tartárico, ácido fosfórico y similares.

65 Ejemplos de vehículos acuosos y no acuosos adecuados que pueden emplearse en las composiciones farmacéuticas de la invención incluyen agua, etanol, polioles (tales como glicerol, propilenglicol, polietilenglicol y similares), y mezclas adecuadas de los mismos, aceites vegetales tales como aceite de oliva, y ésteres orgánicos inyectables tales como oleato de etilo. La fluidez apropiada puede mantenerse, por ejemplo, por el uso de materiales de recubrimiento tales como lecitina, por el mantenimiento del tamaño de partícula requerido en el caso de

dispersiones, y por el uso de tensioactivos.

Estas composiciones también pueden contener adyuvantes tales como conservantes, humectantes, emulsionantes y dispersantes. La prevención de la presencia de microorganismos puede garantizarse tanto por procedimientos de esterilización como por la inclusión de diversos agentes antibacterianos y antifúngicos, por ejemplo, parabeno, clorobutanol, ácido fenolsorbico y similares. También puede desearse incluir agentes isotónicos tales como azúcares, cloruro sódico y similares en las composiciones. Además, la absorción prolongada de la forma farmacéutica inyectable puede provocarse por la inclusión de agentes que retrasan la absorción tal como monoestearato de aluminio y gelatina.

Vehículos farmacéuticamente aceptables incluyen disoluciones o dispersiones acuosas estériles y polvos estériles para la preparación extemporánea de disoluciones inyectables estériles o dispersión. El uso de tales medios y agentes para sustancias farmacéuticamente activas se conoce en la técnica. Excepto en tanto que cualquier medio o agente convencional sea incompatible con el compuesto activo, se contempla el uso del mismo en las composiciones farmacéuticas de la invención. Compuestos activos suplementarios también pueden incorporarse en las composiciones.

Formulaciones a modo de ejemplo comprenden al menos una variante de proteína de la invención y pueden comprender menores concentraciones de agentes estabilizantes (o disgregantes) que pueden usarse, además de los procedimientos desvelados en el presente documento, para prevenir o reducir la agregación de una proteína. Por consiguiente, procedimientos convencionales usados para prevenir la agregación pueden emplearse en el desarrollo de composiciones farmacéuticas que contienen variantes de proteína producidas mediante los procedimientos de la presente invención. Por ejemplo, una variedad de compuestos estabilizantes o disgregantes puede incluirse en composiciones farmacéuticas de la invención dependiendo de su uso previsto y su toxicidad biológica. Tales compuestos estabilizantes pueden incluir, por ejemplo, ciclodextrina y sus derivados (patente de EE.UU. nº 5730969), composiciones de alquilglucósido (solicitud de patente de EE.UU. nº 11/474.049), el uso de moléculas de chaperona (por ejemplo, LEA (Goyal y col., Biochem J. 2005, 388(Pt 1):151-7; los procedimientos de la patente de EE.UU. nº 5688651), compuestos de betaina (Xiao, Burn, Tolbert, Bioconj Chem. 23 de mayo de 2008), tensioactivos (por ejemplo, Pluronic F127, Pluronic F68, Tween 20 (Wei y col. International Journal of Pharmaceutics. 2007, 338(1-2):125-132)), y los procedimientos descritos en las patentes de EE.UU. nº 5696090, 5688651 y 6420122.

Formulaciones a modo de ejemplo también comprenden una variante de proteína de la invención que presenta una propensión alterada a la interacción con un componente de unión junto con un vehículo, adyuvante y/o excipiente farmacéuticamente aceptable.

Además, las proteínas, y en particular los anticuerpos, se estabilizan en formulaciones usando combinaciones de diferentes clases de excipientes, por ejemplo, (1) los disacáridos (por ejemplo, sacarosa, trehalosa) o polioles (por ejemplo, sorbitol, manitol) actúan de estabilizadores por exclusión preferencial y también pueden actuar de crioprotectores durante la liofilización, (2) los tensioactivos (por ejemplo, polisorbato 80, polisorbato 20) actúan minimizando las interacciones de proteínas sobre interfases como líquido/hielo, líquido/superficie de material y/o interfases líquido/aire y (3) los tampones (por ejemplo, fosfato, citrato, histidina) ayudan a controlar y mantener el pH de la formulación. Por consiguiente, pueden usarse tales disacáridos, polioles, tensioactivos y tampones, además de los procedimientos de la presente invención, para estabilizar adicionalmente proteínas y prevenir su agregación.

Las composiciones terapéuticas normalmente deben ser estériles y estables en las condiciones de fabricación y almacenamiento. La composición puede formularse como una disolución, microemulsión, liposoma u otra estructura ordenada adecuada para alta concentración de fármaco. El vehículo puede ser un disolvente o medio de dispersión que contiene, por ejemplo, agua, etanol, poliol (por ejemplo, glicerol, propilenglicol y polietilenglicol líquido, y similares), y mezclas adecuadas de los mismos. La fluidez apropiada puede mantenerse, por ejemplo, por el uso de un recubrimiento tal como lecitina, por el mantenimiento del tamaño de partícula requerido en el caso de la dispersión y por el uso de tensioactivos. En muchos casos será preferible incluir agentes isotónicos, por ejemplo, azúcares, polialcoholes tales como manitol, sorbitol o cloruro sódico en la composición. La absorción prolongada de las composiciones inyectables puede provocarse incluyendo en la composición un agente que retrasa la absorción, por ejemplo, sales de monoestearato y gelatina.

Pueden prepararse disoluciones inyectables estériles incorporando el compuesto activo en la cantidad requerida en un disolvente apropiado con uno o una combinación de los componentes enumerados anteriormente, según se requiera, seguido de microfiltración por esterilización. Generalmente, las dispersiones se preparan incorporando el compuesto activo en un vehículo estéril que contiene un medio de dispersión básico y los otros componentes requeridos de aquellos enumerados anteriormente. En el caso de polvos estériles para la preparación de disoluciones inyectables estériles, los procedimientos preferidos de preparación son secado a vacío y secado por congelación (liofilización) que dan un polvo del principio activo más cualquier componente deseado adicional de una disolución previamente esterilizada por filtración del mismo.

La cantidad de principio activo que puede combinarse con un material de vehículo para producir una única forma de

dosificación variará dependiendo del sujeto que está tratándose, y el modo particular de administración. La cantidad de principio activo que puede combinarse con un material de vehículo para producir una única forma de dosificación será generalmente la cantidad de la composición que produzca un efecto terapéutico. Generalmente, de cada cien por cien, esta cantidad oscilará de aproximadamente el 0,01 por ciento a aproximadamente el noventa y nueve por ciento de principio activo, preferentemente de aproximadamente el 0,1 por ciento a aproximadamente el 70 por ciento, lo más preferentemente de aproximadamente el 1 por ciento a aproximadamente el 30 por ciento de principio activo en combinación con un vehículo farmacéuticamente aceptable.

Las pautas de dosificación se ajustan para proporcionar la respuesta deseada óptima (por ejemplo, una respuesta terapéutica). Por ejemplo, puede administrarse un único bolo, pueden administrarse varias dosis divididas con el tiempo o la dosis puede reducirse o aumentarse proporcionalmente como se indica por las exigencias de la situación terapéutica. Es especialmente ventajoso formular composiciones parenterales en forma unitaria de dosificación para facilitar la administración y uniformidad de la dosificación. Forma unitaria de dosificación como se usa en el presente documento se refiere a unidades físicamente discretas adecuadas como dosificaciones unitarias para los sujetos que van a tratarse; cada unidad contiene una cantidad predeterminada de compuesto activo calculada para producir el efecto terapéutico deseado en asociación con el vehículo farmacéutico requerido. La especificación para las formas unitarias de dosificación de la invención está dictada por y depende directamente de (a) las características únicas del compuesto activo y el efecto terapéutico particular a lograr, y (b) las limitaciones inherentes en la materia de la combinación de un compuesto activo tal para el tratamiento de sensibilidad en individuos.

Para la administración de la proteína, la dosificación oscila de aproximadamente 0,0001 a 100 mg/kg, y más normalmente 0,01 a 5 mg/kg, del peso corporal del huésped. Por ejemplo, las dosificaciones pueden ser 0,3 mg/kg de peso corporal, 1 mg/kg de peso corporal, 3 mg/kg de peso corporal, 5 mg/kg de peso corporal o 10 mg/kg de peso corporal, o dentro del intervalo de 1-10 mg/kg. Una pauta de tratamiento a modo de ejemplo implica administración de una vez por semana, una vez cada dos semanas, una vez cada tres semanas, una vez cada cuatro semanas, una vez al mes, una vez cada 3 meses o una vez cada tres a 6 meses. Pautas de dosificación preferidas para una proteína de la invención incluyen 1 mg/kg de peso corporal o 3 mg/kg de peso corporal mediante administración intravenosa, administrándose el anticuerpo usando uno de los siguientes programas de dosificación: (i) cada cuatro semanas para seis dosificaciones, luego cada tres meses; (ii) cada tres semanas; (iii) 3 mg/kg de peso corporal una vez seguido de 1 mg/kg de peso corporal cada tres semanas.

Alternativamente, una proteína preparada usando un procedimiento de la invención puede administrarse como una formulación de liberación sostenida, en cuyo caso se requiere administración menos frecuente. La dosificación y frecuencia varían dependiendo de la semivida de la sustancia administrada en el paciente. En general, anticuerpos humanos muestran semivida más larga, seguido de anticuerpos humanizados, anticuerpos quiméricos y anticuerpos no humanos. La dosificación y frecuencia de administración pueden variar dependiendo de si el tratamiento es profiláctico o terapéutico. En aplicaciones profilácticas, una dosificación relativamente baja se administra a intervalos relativamente poco frecuentes durante un largo periodo de tiempo. Algunos pacientes continúan recibiendo el tratamiento durante el resto de sus vidas. En aplicaciones terapéuticas, una dosificación relativamente alta a intervalos relativamente cortos se requiere algunas veces hasta que se reduzca o termine la progresión de la enfermedad, y preferentemente hasta que el paciente muestre mejora parcial o completa de los síntomas de enfermedad. Después, al paciente puede administrársele una pauta profiláctica.

Los actuales niveles de dosificación de los principios activos en las composiciones farmacéuticas de la presente invención pueden variarse de forma que se obtenga una cantidad de principio activo que es eficaz para lograr la respuesta terapéutica deseada para un paciente particular, composición y modo de administración, sin ser tóxico para el paciente. El nivel de dosificación seleccionado dependerá de una variedad de factores farmacocinéticos que incluyen la actividad de las composiciones particulares de la presente invención empleadas, o el éster, sal o amida de las mismas, la vía de administración, el tiempo de administración, la tasa de secreción del compuesto particular que se emplea, la duración del tratamiento, otros fármacos, compuestos y/o materiales usados en combinación con las composiciones particulares empleadas, la edad, sexo, peso, condición, salud general e historia médica anterior del paciente que está tratándose, y factores similares muy conocidos en las ciencias médicas.

Una "dosificación terapéuticamente eficaz" de proteína de la invención produce preferentemente una disminución en la gravedad de los síntomas enfermedad, un aumento en la frecuencia y duración de periodos libres de síntomas de enfermedad, o una prevención del deterioro o incapacidad debido a la afección de la enfermedad. Por ejemplo, para el tratamiento de tumores, una "dosificación terapéuticamente eficaz" inhibe preferentemente el crecimiento celular o crecimiento tumoral al menos aproximadamente el 20%, más preferentemente al menos aproximadamente el 40%, incluso más preferentemente al menos aproximadamente el 60%, y todavía más preferentemente al menos aproximadamente el 80% con respecto a sujetos sin tratar. La capacidad de un compuesto para inhibir el crecimiento tumoral puede evaluarse en un sistema de modelo animal predictivo de la eficacia en tumores humanos. Alternativamente, esta propiedad de una composición puede evaluarse examinando la capacidad del compuesto para inhibir tal inhibición *in vitro* por ensayos conocidos para el médico habitual. Una cantidad terapéuticamente eficaz de un compuesto terapéutico puede disminuir el tamaño del tumor, o de otro modo mejorar los síntomas en un sujeto. Un experto habitual en la materia podría determinar tales cantidades basándose en factores tales como el tamaño del sujeto, la gravedad de los síntomas del sujeto y la composición particular o vía de administración

seleccionada.

Una composición de la presente invención puede administrarse mediante una o más vías de administración usando uno o más de una variedad de procedimientos conocidos en la técnica. Como será apreciado por el experto, la vía y/o modo de administración variará dependiendo de los resultados deseados. Vías de administración preferidas para unir restos de la invención incluyen vías de administración intravenosa, intramuscular, intradérmica, intraperitoneal, subcutánea, espinal u otras vías de administración parenteral, por ejemplo, mediante inyección o infusión. El término "administración parenteral" como se usa en el presente documento significa modos de administración distintos de administración enteral y tópica, normalmente mediante inyección, e incluyen, sin limitación, inyección e infusión intravenosa, intramuscular, intra-arterial, intratecal, intracapsular, intraorbital, intracardiaca, intradérmica, intraperitoneal, transtraqueal, subcutánea, subcuticular, intraarticular, subcapsular, subaracnoidea, intraespinal, epidural e intraesternal.

Alternativamente, una proteína preparada usando un procedimiento de la invención puede administrarse mediante una vía no parenteral tal como una vía de administración tópica, epidérmica o mucosa, por ejemplo, intranasalmente, por vía oral, vaginalmente, rectalmente, sublingualmente o tópicamente.

Los compuestos activos pueden prepararse con vehículos que protegerán al compuesto de la rápida liberación, tales como una formulación de liberación controlada, que incluyen implantes, parches transdérmicos y sistemas de administración microencapsulados. Pueden usarse polímeros biocompatibles biodegradables, tales como etilenoacetato de vinilo, polianhídridos, ácido poliglicólico, colágeno, poliortoésteres y ácido poliláctico. Muchos procedimientos para la preparación de tales formulaciones están patentados o generalmente son conocidos para aquellos expertos en la materia. Véase, por ejemplo, *Sustained and Controlled Release Drug Delivery Systems*, J.R. Robinson, ed., Marcel Dekker, Inc., Nueva York, 1978.

Las composiciones terapéuticas pueden administrarse con dispositivos médicos conocidos en la técnica. Por ejemplo, en una realización preferida, una composición terapéutica de la invención puede administrarse con un dispositivo de inyección hipodérmica sin aguja tal como los dispositivos desvelados en las patentes de EE.UU. nº 5.399.163; 5.383.851; 5.312.335; 5.064.413; 4.941.880; 4.790.824; o 4.596.556. Ejemplos de implantes y módulos muy conocidos útiles en la presente invención incluyen: patente de EE.UU. nº 4.487.603, que desvela una bomba de micro-infusión implantable para dispensar medicación a una tasa controlada; la patente de EE.UU. nº 4.486.194, que desvela un dispositivo terapéutico para administrar medicamentos a través de la piel; la patente de EE.UU. nº 4.447.233, que desvela una bomba de infusión de medicación para administrar medicación a una tasa de infusión precisa; la patente de EE.UU. nº 4.447.224, que desvela un aparato de infusión implantable de flujo variable para la administración de fármaco continua; la patente de EE.UU. nº 4.439.196, que desvela un sistema de administración de fármaco osmótico que tiene compartimentos de múltiples cámaras; y la patente de EE.UU. nº 4.475.196, que desvela un sistema de administración de fármaco osmótico. Muchos otros de tales implantes, sistemas de administración y módulos son conocidos para aquellos expertos en la materia.

EJEMPLOS

Introducción a los ejemplos

Las técnicas de simulación molecular para predecir regiones propensas a la agregación y estudiar el mecanismo de agregación han empleado generalmente modelos de simulación comparativamente simples (Ma y Nussinov. *Curr. Opin. Chem. Biol.* 2006, 10, 445-452; Cellmer y col., *TRENDS in Biotechnology* 2007, 25(6), 254) a diferencia de los modelos atomísticos detallados que pueden emplearse en la presente invención. El menos detallado de los modelos de simulación empleados fue el modelo de red cristalina, que se usó en numerosos estudios de agregación de proteínas (Harrison y col. *J. Mol. Biol.* 1999, 286,593-606; Dima y Thirumalai. *Protein Sci.* 2002, 11, 1036-1049; Leonhard y col. *Protein Sci.* 2004, 13, 358-369; Patro y Przybycien. *Biophys. J.* 1994, 66, 1274-1289; Patro y Przybycien. *Biophys. J.* 1996, 70, 2888-2902; Broglia y col. *Proc. Natl. Acad. Sci. U.S.A.* 1998, 95, 12930-12933; Istrail y col. *Comput. Biol.* 1999, 6, 143-162; Giugliarelli y col. *Chem. Phys.* 2000, 113, 5072-5077; Bratko y col. *J. Chem. Phys.* 2001, 114,561-569; Bratko y Blanch J. *Chem. Phys.* 2003, 118, 5185-5194; Combe y Frenkel *Chem. Phys.* 2003, 118, 9015-9022; Toma y Toma. *Biomacromolecules* 2000, 1, 232-238; Gupta y col. *Protein Sci.* 1998, 7, 2642-2652; y Nguyen y Hall *Biotechnol. Bioeng.* 2002, 80, 823-834). Aquí, cada residuo se representa como una perla que ocupa un único sitio sobre una red cristalina tridimensional. Debido a su simplicidad, el modelo de red cristalina es menos exigente computacionalmente y se ha usado para simular grandes sistemas para escalas a largo plazo. Aunque estos modelos de la red cristalina proporcionan conocimiento en la física básica subyacente a la agregación de proteínas, no representan con exactitud la estructura secundaria y terciaria, y no pueden explicar adecuadamente diferentes interacciones a nivel atomístico tales como el enlace de hidrógeno.

Un modelo más detallado en comparación con el modelo de red cristalina es el modelo de resolución intermedia en el que algunos átomos se combinan normalmente en una única perla, y algunas veces se introducen pseudo-enlaces para mantener los ángulos de enlace del esqueleto y los estados de isomerización (Smith y Hall, *Mol. Biol.* 2001, 312, 187-202; Smith y Hall. *Proteins: Struct., Funct., Genet.* 2001, 44, 344-360; Smith y Hall. *Proteins: Struct., Funct., Genet.* 2001, 44, 376-391; Nguyen y col., *Protein Sci.* 2004, 13, 2909-2924; Nguyen y Hall, *Proc. Natl. Acad. Sci.*

- U.S.A., 2004, 101(46), 16180-16185; Nguyen y Hall. J. Am. Chem. Soc., 2006, 128, 1890-1901; Jang y col., Biophys. J. 2004, 86, 31-49; Jang y col., Protein Sci. 2004, 13, 40-53). Este modelo se usó satisfactoriamente para simular la formación de fibrillas de sistemas que contenían entre 12 y 96 péptidos de polialanina (16 residuos cada uno) a partir de un estado aleatorio (Nguyen y Hall, Proc. Natl. Acad. Sci. U.S.A., 2004, 101(46), 16180-16185; Nguyen y Hall, J. Am. Chem. Soc., 2006, 128, 1890-1901). Dokholyan y colaboradores aplicaron un modelo tal para estudiar la formación de estructuras de hoja β fibrilares por ocho proteínas del dominio Src SH3 del modelo (Ding y col., Mol. Biol. 2002, 324, 851-857) o por 28 péptidos de A β (1-40) del modelo (Peng y col., Fis. Rev. E: Stat. Ph. Interdiscip. Top. 2004, 69, 41908-41914.).
- 5 A diferencia de modelos más simples, los modelos atomísticos incluyen todos los detalles atomísticos tales como enlace de hidrógeno y así son más exactos que los modelos de la red cristalina o de resolución intermedia. Tales modelos atomísticos se han usado tanto con un disolvente explícito como con un disolvente implícito en los que el disolvente se trata como un continuo. El modelo explícito es más exacto que el modelo implícito, pero también es más exigente computacionalmente. Un modelo atomístico tal con disolvente implícito se usó para estudiar las fases tempranas de agregación del heptapéptido GNNQQNY (SEC ID N°: 1), que es una parte de la proteína de la levadura Sup35 (Gsponer y col., Proc. Natl. Acad. Sci. U.S.A. 2003, 100, 5154-5159). Se usó un modelo similar para la agregación del péptido amiloide Ab16-22 (KLVFFAE (SEC ID N°: 2) en hojas β antiparalelas (Klimov y Thirumalai, Structure 2003, 11, 295-307). Dokholyan y colaboradores (Khare y col., Proteins. 2005, 61, 617-632) usaron un modelo atomístico explícito para investigar la propensión a la agregación ordenada a lo largo de la secuencia de la enzima superóxido dismutasa de Cu, Zn (SOD 1). Han descompuesto la secuencia de SOD1 en heptapéptidos solapantes y realizado un gran número de simulaciones de dinámica molecular en agua explícitas (cada una de 0,5 ns) de segmentos monoméricos, diméricos y tetaméricos. Con esto identificaron que las regiones amiloidogénicas en la secuencia de SOD1 eran: los dos extremos, las cadenas β 4 y 7 y los dos bucles cruzados.
- 10 Se desarrolló un protocolo de simulación de dinámica molecular similar para obtener información estructural sobre la agregación β ordenada de polipéptidos amiloidogénicos (Cecchini y col., J Mol Biol. 2006, 357, 1306-1321). El procedimiento se basa en la descomposición de una cadena de polipéptidos en segmentos solapantes y las simulaciones de dinámica molecular (MD) en equilibrio de un número pequeño de copias de cada segmento. La propensión a la agregación β a lo largo de la secuencia del péptido A β (1-42) de Alzheimer se encontró que era altamente heterogénea con un máximo en el segmento V₁₂HHQKLVFFAA₂₂ (SEC ID N°: 3) y un mínimo en cuatro dipéptidos similares a giro. Usando esta técnica, el cambio predicho en la propensión a la agregación de un mutante de punto doble del dominio del extremo N del prión de la levadura Ura2p se verificó *in vitro* usando el ensayo de unión a tioflavina T. Un procedimiento tal para descomponer la cadena de polipéptidos en segmentos solapantes sería extremadamente desafiante para sistemas tales como anticuerpos debido a su gran tamaño. Incluso una simulación atomística de un único anticuerpo completo en disolvente explícito es muy exigente computacionalmente debido al enorme tamaño de un anticuerpo. Por tanto, no parece que sea simulación atomística de anticuerpos completos en la bibliografía.
- 25 Sin embargo, ha habido simulaciones atomísticas de partes pequeñas del anticuerpo, generalmente para el fragmento Fab (Noon y col., PNAS. 2002, 99, 6466; Sinha y Smith-Gill, Cell Biochemistry and Biophysics. 2005, 43, 253). En el presente trabajo se realizaron simulaciones atomísticas de una molécula de anticuerpo completa con un disolvente explícito. Basándose en estas simulaciones, las regiones propensas a la agregación sobre el anticuerpo se identificaron usando el parámetro 'propensión a la agregación espacial' descrito en el presente documento. Estas regiones propensas a la agregación se mutaron entonces para diseñar anticuerpos con estabilidad potenciada. Los ejemplos descritos en el presente documento se refieren a realizaciones particulares de la invención.
- 30
- 35
- 40
- 45

Ejemplo 1: Metodología de simulación de dinámica molecular

- Se realizaron simulaciones de dinámica molecular para un anticuerpo completo usando un modelo de todos los átomos. La estructura inicial para la simulación para el anticuerpo completo se obtuvo de las estructuras de rayos X de fragmentos Fab y Fc individuales. La estructura de rayos X de un fragmento Fab de prueba de concepto (POC) se seleccionó para modelar sobre la estructura de rayos X de Fc obtenida a partir del anticuerpo IgG1 1HZH (Saphire y col., Science. 2001, 293, 1155). 1HZH se eligió ya que la estructura de rayos X es conocida para el anticuerpo completo y ya que la estructura de Fc es la misma para todos los de la clase IgG1 de anticuerpos. La estructura de un anticuerpo POC completo se obtuvo entonces alineando los fragmentos Fab y Fc usando la estructura de 1HZH como molde del modelo. Con el fin de alinear los fragmentos a la distancia y orientación correcta, la RMSD (desviación de la raíz cuadrada de la media) se minimizó entre los residuos CYS comunes de los fragmentos y el molde del anticuerpo completo (1HZH). Los residuos CYS se eligieron debido a que cada subdominio de anticuerpo (cH1, cH2 etc.) contiene un enlace disulfuro, y así, residuos CYS están ampliamente distribuidos a través de la estructura del anticuerpo completo. La estructura del anticuerpo completo resultante se usó entonces para realizar simulaciones de átomos explícitos durante 30 ns. Se usó un patrón de glucosilación G0 para las simulaciones ya que éste es el patrón de glucosilación más común observado en anticuerpos.
- 50
- 55
- 60

- El paquete de simulación CHARMM (Brooks y col. J. Comput. Chem., 1983, 4, 187) se usó para la configuración y análisis y el paquete NAMD (Phillips y col. Journal of Computational Chemistry. 2005, 26, 1781) para realizar
- 65

simulaciones. El campo de fuerza completamente atómico de CHARMM (MacKerell y col. J. Phys Chem. B. 1998, 102, 3586) se usó para la proteína y el modelo de disolvente TIP3P (Jorgensen y col. J. Chem. Phys., 1983, 79, 926) para agua. Las simulaciones se realizaron a 298K (24,85 °C) y 1 atm (0,1 MPa) en el colectivo NPT. Los parámetros para los grupos de azúcar implicados en la glucosilación del fragmento Fc se derivaron para estar de acuerdo con el campo de fuerza de CHARMM, después del campo de fuerza de CSFF (Kuttel y col. J. Comput. Chem., 2002, 23, 1236). Los estados de protonación de residuos de histidina a pH 7 se eligieron basándose en la proximidad espacial de grupos electro-negativos. El anticuerpo completo se solató en una caja ortorrómbica, ya que ésta minimiza el número de moléculas de agua requeridas y así minimiza el tiempo computacional. Se usaron condiciones límite periódicas en las 3 direcciones. Se usó una vaina de solvatación de agua de 8 Å en cada dirección de la caja ortorrómbica. El tamaño del sistema total resultante fue 202130 átomos. Se añadieron iones suficientes para neutralizar la carga total del sistema. La neutralidad de la carga se requiere por la técnica de las sumas de Ewald empleada para calcular la contribución de interacciones electrostáticas en el sistema.

Después de solvatar el anticuerpo, la energía se minimizó inicialmente con SD (descensos de mayor pendiente) fijando la proteína para permitir que el agua se relajara alrededor de la proteína. Entonces, las restricciones se eliminaron y la estructura se minimizó adicionalmente con SD y ABNR (Newton-Raphson de base adoptada). El sistema se calentó entonces lentamente hasta temperatura ambiente con incrementos de 5 °C cada 0,5 ps usando una etapa de menos tiempo. Entonces, el sistema se equilibró para 1 ns antes de calcular las propiedades de interés de la simulación. Las configuraciones se guardaron cada 0,1 ps durante la simulación para el posterior análisis estadístico.

Ejemplo 2: Cálculo de la propensión a la agregación espacial (SAP)

Con el fin de vencer los fallos de SAA, se definió un nuevo parámetro llamado 'propensión a la agregación espacial' como se ha descrito anteriormente.

En este ejemplo, la 'propensión a la agregación espacial' se calculó para regiones esféricas con radio R centrado en cada átomo en el anticuerpo descrito en el Ejemplo 1. El valor de propensión a la agregación espacial se evaluó así con un promedio de simulación de 30 ns para el fragmento de Fc del anticuerpo para dos radios diferentes de parches (R=5 Å, 10 Å) (un experto en la materia apreciará que pueden elegirse diversas etapas de tiempo para la simulación según los recursos computacionales disponibles y la resolución deseada del resultado). En ambos casos se observó que la mayoría de los valores fueron negativos, que indica que las regiones más expuestas son hidrófilas. Esto era de esperar ya que la mayor parte de la superficie de la proteína expuesta es normalmente hidrófila. También se observó que había algunas regiones con picos positivos para la propensión a la agregación espacial que indicaban alta hidrofobia expuesta. Yendo de radios más pequeños de parches (5 Å) a los radios más grandes (10 Å) se eliminan algunos picos, mientras que algunos otros picos se potencian. Algunos picos se eliminaron debido a que en estas regiones un pequeño parche hidrófobo (con menos de 5 Å de radio) está rodeado de parches hidrófilos; así, promediando sobre 10 Å conduce a una disminución eficaz en la hidrofobia para la región. Mientras, en algunas otras regiones, la propensión a la agregación espacial a R=10 Å se potencia debido a parches hidrófobos que rodean un parche hidrófobo similar.

Anteriormente, la propensión a la agregación espacial se calculó como un promedio durante la ejecución de la simulación de 30 ns. Los resultados calculados usando la simulación se compararon entonces con la propensión a la agregación espacial de solo la estructura de rayos X, sin simulación molecular. La propensión a la agregación espacial (rayos X) fue similar a la del valor promediado para la simulación, que tiene picos en las mismas localizaciones pero con diferencias en la magnitud de los picos. Las diferencias fueron mayores con el mayor radio de parche, R=10 Å. Esto es probablemente debido a que las diferencias son aditivas cuando se mira a tamaños de parches más grandes. Estas diferencias se producen debido a la exposición de superficie cambiante de los residuos en la ejecución de simulación dinámica. Sin embargo, esta comparación muestra que puede obtenerse una buena estimación inicial de la propensión a la agregación espacial, especialmente para bajo radio del parche R, a partir de la propia estructura de rayos X.

Los valores de propensión a la agregación espacial de la simulación para R=5 Å y 10 Å se mapearon sobre la estructura del anticuerpo. En ambos casos, la superficie del anticuerpo se coloreó según los valores de la propensión a la agregación espacial. Valores positivos de la propensión a la agregación espacial (hidrófobos) se muestran en gris o negro mientras que valores negativos (hidrófilos) están en gris más claro o blanco. La intensidad de color es proporcional a la magnitud de SES. Por tanto, un parche hidrófobo altamente expuesto sería negro intenso total, y similarmente uno hidrófilo altamente expuesto será blanco más brillante. La representación estructural del anticuerpo también se basa en el área accesible al disolvente para cada residuo. A ambos radios usados en el cálculo de la propensión a la agregación espacial (5 Å y 10 Å) se observó que la superficie es predominantemente blanca, que indica que la superficie es generalmente hidrófila. Esto es de nuevo según se esperaba, ya que la mayor parte de la superficie de la proteína es normalmente hidrófila. Sin embargo, algunas áreas negras son perceptibles, que indica regiones hidrófobas expuestas. El contraste entre las regiones negras y blancas es más prominente a los mayores radios del parche usados en el cálculo de SAP, R=10 Å. Estas regiones negras (hidrófobas) tienen excelente correlación con regiones del anticuerpo conocidas por interactuar con otras proteínas: una región negra intensa en la región bisagra es en la que el receptor de Fc interactúa, una región

negra en el fragmento Fc es en la que la proteína A y la proteína G interaccionan, y un parche negro al final del fragmento Fab es en el que el anticuerpo se une a antígenos. La propensión a la agregación espacial se representó para $R=5 \text{ \AA}$ y 10 \AA , respectivamente, pudiendo observarse la misma correlación de picos con regiones interactuantes. Los sitios de interacción de proteínas se obtuvieron de la estructura de rayos X de los complejos de proteína, entradas de PDB 1T89, 1FC2 y 1FCC (Radaev, J. Biol. Chem. 2001, 276 (19) 16469; Deisenhofer y col. Hoppe-Seyler's Z Physiol Chem. 1978. 359, 975-985; Deisenhofer, J. Biochemistry. 1981, 20, 2361-2370; Sauer-Eriksson y col. Structure. 1995, 3, 265). Las interacciones hidrófobas se correlacionan muy bien con los picos positivos y las interacciones hidrófilas se correlacionan bien con los picos negativos. Por tanto, el parámetro de propensión a la agregación espacial también puede usarse para predecir los sitios de unión de proteínas. En las pocas excepciones en las que también interaccionan residuos con baja propensión a la agregación espacial (es decir, próximos a cero, tanto positivos como negativos), se observó que las interacciones son en realidad con los átomos de la propia cadena principal del esqueleto, en lugar de con las cadenas laterales.

Aparte de los parches negros que ya se ha mostrado que interaccionan con otras proteínas, se identificaron parches negros adicionales sobre la superficie del anticuerpo. Un parche en el fondo de Fc es significativamente hidrófobo, pero está algo enterrado dentro, con región hidrófila sobre sus límites. Similarmente, dos parches son hidrófobos y están expuestos a disolvente, pero están orientados hacia el interior del anticuerpo. Estos parches todavía podrían estar potencialmente implicados en interacciones con otras proteínas si se exponen debido a cambios conformacionales significativos o desplegamiento del anticuerpo. Todos los parches hidrófobos también podrían observarse en el radio de parche más pequeño ($R=5 \text{ \AA}$), aunque con menos contraste en comparación con el radio de parche mayor ($R=10 \text{ \AA}$).

Los valores de propensión a la agregación espacial (rayos X) que se basan en solo la estructura de rayos X también se mapearon sobre la superficie del anticuerpo, para compararlos con los valores promediados de la simulación. Los parches propensos a la agregación hidrófobos negros son bastante similares entre la propensión a la agregación espacial calculada tanto por simulación como usando solo la estructura de rayos X. Por supuesto, hay algunas diferencias, tales como la intensidad de parches en la región en la que interaccionan la proteína A y G. Sin embargo, esta comparación demuestra que la propensión a la agregación espacial (rayos X) basada en solo la estructura de rayos X puede usarse para obtener una buena descripción de la distribución de parches hidrófobos sobre la superficie. Esto es importante ya que la simulación atómica de un anticuerpo completo es computacionalmente exigente. Para proteínas que carecen de un modelo estructural de rayos X, el mismo parámetro de propensión a la agregación espacial puede aplicarse a la estructura generada mediante modelado por homología o predicción de estructura desde el inicio. Se observó que la estructura de homología era muy similar a la estructura de rayos X, y sus valores de propensión a la agregación espacial también son similares a la estructura de rayos X.

Así, la propensión a la agregación espacial identifica los parches hidrófobos sobre la superficie del anticuerpo. Estos parches podrían exponerse nativamente o exponerse debido a fluctuaciones dinámicas o desplegamiento parcial del anticuerpo. Algunos de estos parches hidrófobos también se correlacionan bien con regiones que interaccionan con otras proteínas. Con el fin de probar si estos parches hidrófobos predichos por propensión a la agregación espacial también participan en la agregación, se realizaron mutaciones en estas regiones específicas para cambiar los residuos hidrófobos a residuos hidrófilos. Los anticuerpos resultantes mostraron menos comportamiento de agregación y estabilidad mejorada. Aparte de identificar residuos propensos a la agregación, también se observó que el procedimiento de SAP identifica correctamente las regiones del anticuerpo propensas a unión con otras proteínas. Por tanto, el procedimiento podría aplicarse ampliamente a todas las proteínas para identificar las regiones propensas a la agregación o regiones de unión con otras proteínas.

Ejemplo 3: Selección de sitios de anticuerpo para manipular la estabilidad

Los sitios que van a manipularse para estabilidad del anticuerpo potenciada se seleccionaron basándose en el parámetro de SAP. Este parámetro espacial explica (1) área accesible al disolvente (SAA) de cada residuo, (2) la hidrofobia del residuo, y (3) las contribuciones espaciales de todos los residuos dentro de un cierto radio. En este ejemplo, los residuos hidrófobos que se corresponden con los picos positivos en C_{H2} se cambiaron a residuos no hidrófobos. Se esperaba que esto mejorara la estabilidad global de proteínas. Los dos sitios seleccionados (A1 y A2) se corresponden con dos residuos muy hidrófobos. Se realizó un análisis de sustituciones de estos residuos con lisina, un aminoácido muy hidrófilo con una cadena lateral positivamente cargada. La variante A1 y la variante A2 se diferencian de la natural por la sustitución de un único aminoácido.

Ejemplo 4: Expresión y purificación de las variantes de anticuerpo

Las variantes de anticuerpo se generaron por mutagénesis dirigida a sitio. Todas las construcciones se confirmaron por secuenciación de ADN. El ADN de plásmido a la escala de mg se purificó a partir de cultivos bacterianos y se transfectó transitoriamente en células HEK 293. Anticuerpo natural y variantes se purificaron a partir del sobrenadante de cultivo de tejido sobre una columna de proteína A y se pasaron sobre una columna de Q Sepharose para eliminar impurezas negativamente cargadas. A pH 7,0 e inferior, los anticuerpos están positivamente cargados y siguen en el flujo continuo, mientras que impurezas negativamente cargadas se unen a la matriz positivamente cargada de la columna de Q Sepharose. La disolución con anticuerpo purificado se concentró y

se intercambió el tampón con tampón His 20 mM a pH 6,5 a una concentración final de 150 mg/ml.

Como control de calidad, alícuotas de las muestras purificadas y concentradas se analizaron por SDS-PAGE y dicroísmo circular. Se usaron condiciones tanto reductoras como no reductoras para los geles de proteína. Los presentes inventores también compararon la estructura secundaria del anticuerpo natural y la variante A1 por dicroísmo circular.

Ejemplo 5: Caracterización biofísica

La estabilidad de la variante A1 se comparó con natural en un experimento de agregación acelerada. Muestras a 150 mg/ml en tampón His 20 mM a pH 6,5 se incubaron a 58 °C durante hasta 24 horas. La incubación se detuvo diluyendo la muestra con 10 mg/ml con tampón fosfato de K 15 mM, pH 6,5, y el porcentaje de agregación se determinó por SEC-HPLC. La agregación se calculó como la suma de las áreas de todos los picos no monoméricos dividida entre el área total de todos los picos. Se muestra el promedio de 2-4 muestras para cada momento de tiempo. Los agregados para la variante A1 son tan solo el 80% de los agregados para natural. Así, una única mutación puntual reduce la formación de agregados el 20%.

Se compararon natural y variante A1 por microcalorimetría diferencial de barrido (DSC, Microcal). Anticuerpos completos son proteínas multi-dominio. El análisis de DSC indica diferentes temperaturas de fusión para diferentes dominios (Ionescu, R.M. y col., J Pharm Sci. 2008, 97(4): pág. 1414-26; Mimura, Y. y col., J Biol Chem. 2001, 276(49): pág. 45539-47.). Los dominios C_{H2} y C_{H3} constantes de Fc de IgG1 humana tienen temperaturas de fusión de aproximadamente 70 °C y 82 °C, respectivamente, a pH neutro (Ionescu, R.M. y col., J Pharm Sci. 2008, 97(4): pág. 1414-26; Mimura, Y. y col., Role of oligosaccharide residues of IgG1-Fc in Fc gamma RIIB binding. J Biol Chem, 2001. 276(49): pág. 45539-47). Dependiendo de la secuencia de dominios variables del anticuerpo, los fragmentos Fab pueden tener diferentes temperaturas de fusión con respecto a C_{H2} y C_{H3}. El anticuerpo C contiene un dominio de Fab con transición de desplegamiento que se encuentra entre las transiciones de C_{H2} y C_{H3}. Así, C_{H2} es el dominio de anticuerpo con la menor temperatura de fusión.

Se analizaron natural y variante A1 a una concentración de 2 mg/ml en tampón His 15 mM a pH 6,5 y una tasa de calentamiento de 1,5 grados por minuto. Los datos de la muestra se analizaron por resta de los datos de referencia, normalización a la concentración de proteína y volumen de celda de DSC e interpolación de un nivel inicial cúbico. Una comparación de los termogramas muestra un aumento de la transición de fusión de C_{H2} en la variante A1 en comparación con natural.

El análisis de la variante A2, también manipulada para estabilidad basándose en los valores de propensión a la agregación espacial, recapitula los hallazgos para la variante A1.

En resumen, los análisis biofísicos de las variantes de anticuerpo manipuladas demostraron una agregación reducida y una estabilidad potenciada. La fuerte correlación entre sitios manipulados, estabilidad de variantes y perfiles de DSC es una prueba de la eficacia de la metodología para estabilizar proteínas terapéuticas.

Ejemplo 6: SAA eficaz

Se ha observado que los picos en SAA eficaz (suma de 3 residuos) pueden correlacionarse con regiones propensas a la agregación en una estructura de proteína. Por consiguiente, el SAA eficaz puede usarse como procedimiento separado, aunque menos poderoso, para identificar regiones propensas a la agregación de una proteína. Altos valores de SAA eficaz (suma de 3 residuos) indican las regiones más hidrófobas y bajos valores indican las regiones más hidrófilas. Los datos sobre una proteína de prueba que tiene una tendencia a la formación de agregados se obtuvieron de simulaciones moleculares cortas de 1,2 ns (plegadas) y 1 ns (plegadas erróneamente). El SAA eficaz se representó para residuos de la proteína y se observó que había una buena correlación entre los picos de SAA eficaz y desapareamiento en la red de enlace de la estructura de proteína. Esto indica que el SAA eficaz estuvo identificando con exactitud residuos de la estructura de proteína que fomentan el plegamiento erróneo o la agregación de proteínas. Se hicieron varios mutantes de la proteína de prueba y al menos uno mostró resultados prometedores en retener una estructura de proteína adecuadamente plegada.

Ejemplo 7: Predicción de regiones de unión a proteína usando SAP

Se usó el procedimiento de SAP para predecir sitios de unión a proteína. Se predijeron regiones de unión para dos proteínas diferentes: un anticuerpo IgG1 y EGFR. Un anticuerpo IgG1 es muy conocido por unirse con proteínas tales como receptor de Fc, proteína A y proteína G. El EGFR se une con factor de crecimiento epidérmico (EGF), factor de crecimiento transformante (TGF α) y también consigo mismo para formar un dímero. Estas regiones de unión para anticuerpo IgG1 y EGFR se usaron como modelos para demostrar la capacidad de la herramienta de SAP en predecir regiones de unión.

Procedimientos de simulación molecular

Se realizaron simulaciones de dinámica molecular para un anticuerpo IgG1 completo usando un modelo de todos los átomos con disolvente explícito. La estructura de partida para la simulación se obtuvo uniendo las estructuras de rayos X de fragmentos Fab y Fc individuales del anticuerpo. La estructura de rayos X del fragmento Fab se obtuvo de Novartis Pharma AG. La estructura de rayos X del fragmento Fc se obtuvo a partir de otro anticuerpo IgG1 de secuencia similar, 1HZH (Saphire y col., Science. 2001, 293, 1155). La estructura de un anticuerpo completo se obtuvo entonces alineando los fragmentos Fab y Fc usando la estructura de 1HZH como molde del modelo. Esta estructura del anticuerpo se llamó anticuerpo A. Con el fin de alinear los fragmentos a la distancia y orientación correcta, la RMSD (desviación de la raíz cuadrada de la media) se minimizó entre los residuos CYS comunes de los fragmentos y el molde del anticuerpo completo (1HZH). Entonces, esta estructura se usó para realizar simulaciones de átomos explícitos durante 30 ns. Los residuos CYS en el anticuerpo A resultante participaron todos en enlaces disulfuro, que incluyen aquellos en la región bisagra. Se usó un patrón de glucosilación G0 para las simulaciones, ya que éste es uno de los patrones de glucosilación más comunes observados en anticuerpos.

El paquete de simulación CHARMM (Brooks y col. J. Comput. Chem., 1983, 4, 187) se usó para la configuración y análisis y el paquete NAMD (Phillips y col. Journal of Computational Chemistry. 2005, 26, 1781) para realizar simulaciones. El campo de fuerza completamente atómica de CHARMM (Phillips y col. Journal of Computational Chemistry. 2005, 26, 1781) se usó para la proteína y el modelo de disolvente TIP3P (Jorgensen y col. J. Chem. Phys., 1983, 79, 926) para agua. Las simulaciones se realizaron a 298K (24,85 °C) y 1 atm (0,1 MPa) en el colectivo NPT. Los parámetros para los grupos de azúcar implicados en la glucosilación del fragmento Fc se derivaron en coherencia con el campo de fuerza de CHARMM, después del campo de fuerza de CSFF (Kuttel y col. J. Comput. Chem., 2002, 23, 1236). Los estados de protonación de residuos de histidina a pH 7 se decidieron basándose en la proximidad espacial de grupos electro-negativos. El anticuerpo completo se solvató en una caja ortorrómbica, ya que ésta minimiza el número de moléculas de agua requeridas y así minimiza el tiempo computacional requerido. Se usaron condiciones límite periódicas en las 3 direcciones. Se usó una vaina de solvatación de agua de 8 Å en cada dirección de la caja ortorrómbica. El tamaño del sistema total resultante fue 202.130 átomos. Se observó que la caja ortorrómbica permaneció estable durante la simulación de 30 ns sin ningún cambio significativo en las dimensiones de la caja en los tres ejes. Las dimensiones iniciales de la caja fueron 161,9 Å, 145,4 Å y 83,2 Å, respectivamente, y cambiaron muy poco durante la simulación de 30 ns, terminando en 161,2 Å, 144,7 Å y 82,8 Å, respectivamente. El anticuerpo no giró significativamente durante la simulación de 30 ns, manteniéndose así la distancia mínima entre el anticuerpo y sus imágenes periódicas de más de 14 Å. Se añadieron iones suficientes para neutralizar la carga total del sistema. La neutralidad de la carga se requiere por la técnica de las sumas de Ewald empleada para calcular la contribución debida a las interacciones electrostáticas.

Después de solvatar el anticuerpo, la energía se minimizó inicialmente con SD (descensos de mayor pendiente) fijando la proteína para permitir que el agua se relajara alrededor de la proteína. Entonces, las restricciones se eliminaron y la estructura se minimizó adicionalmente con SD y ABNR (Newton-Raphson de base adoptada). El sistema se calentó entonces lentamente hasta temperatura ambiente con incrementos de 5 °C cada 0,5 ps usando una etapa de 1 fs de tiempo. Entonces, el sistema se equilibró para 1 ns antes de empezar el cálculo de las diversas propiedades de la simulación. Las configuraciones se guardaron cada 0,1 ps durante la simulación para el posterior análisis estadístico.

Herramienta de SAP para predecir regiones de unión de un anticuerpo IgG1

La herramienta de SAP se aplicó a las configuraciones de proteína obtenidas de simulaciones moleculares. Para predicciones más rápidas en aplicaciones de alta resolución, la herramienta de SAP también puede aplicarse a la estructura de rayos X de proteínas o estructura derivada por homología, con una advertencia de que podría conducir a una pérdida de precisión. El valor de SAP para cada átomo en la proteína se definió del siguiente modo,

$$\left(\text{Pr opensión a la agregación} \right)_{\text{espacial (SAP)}} = \sum_{\text{átomo } i} \left\{ \sum_{\substack{\text{Residuos con al menos un} \\ \text{átomo de la cadena lateral} \\ \text{dentro de R del átomo } i}} \left(\frac{\text{SAA de átomos de la cadena lateral dentro del radio R}}{\text{SAA de átomos de la cadena lateral del residuo completamente expuesto}} \times \text{Hidrofobia del átomo} \right) \right\}$$

Aquí,

- 1) SAA de átomos de la cadena lateral dentro del radio R se calcula en cada instante de la simulación
- 2) SAA de la cadena lateral del residuo completamente expuesto (dicho para el aminoácido 'X') se obtiene calculando el SAA de cadenas laterales del residuo central en la conformación completamente extendida de triptido 'Ala-X-Ala'.

3) Hidrofobia del residuo se obtiene a partir de la escala de hidrofobia de Black y Mould (S. D. Black y D. R. Mould, Anal. Biochem. 193, 72 (1991)). La escala se normaliza de forma que la glicina tenga una hidrofobia de cero. Por tanto, los aminoácidos que son más hidrófobos que la glicina son positivos y los menos hidrófobos que la glicina son negativos en la escala hidrófoba.

5 SAP da la hidrofobia dinámicamente expuesto de un cierto parche centrado en el átomo dado sobre la superficie de la proteína. SAP se calcula para regiones esféricas con el radio R centrado en cada átomo en la proteína. Esto da un único valor de SAP para cada átomo. Entonces, la SAP para un residuo se obtiene promediando la SAP de todos sus átomos constituyentes. Los valores de SAP se evaluaron así usando $R=10 \text{ \AA}$ para un anticuerpo IgG1, y los valores se mapearon sobre la superficie del anticuerpo usando una escala de color para indicar el valor de SAP dentro de un intervalo de -0,5 a +0,5. Estos valores de SAP se calcularon promediando sobre la simulación atomística del anticuerpo completo de 30 ns. Obsérvese que el valor de SAP en cada residuo da la hidrofobia expuesta total de un parche centrada en esa residuo, y no solo la hidrofobia para un único residuo. La escala de hidrofobia (S. D. Black y D. R. Mould, Anal. Biochem. 193, 72 (1991)) también se mapeó directamente sobre la superficie para comparación. Si se visualiza en mapa hidrófobo, las regiones hidrófobas parecieren estar distribuidas aleatoriamente por toda la superficie, y sería difícil escoger una cierta región hidrófoba que fuera más dominante en comparación con la otra. Sin embargo, tras examinar el mapa de SAP de la misma estructura, fue fácil localizar las regiones de alta SAP, que indican regiones hidrófobas dinámicamente expuestas. Es termodinámicamente desfavorable para estos parches exponerse a agua debido a su naturaleza hidrófoba. Por tanto, podrían participar en la unión de proteínas con el fin de reducir su exposición a disolvente. Estas regiones de alta SAP se identificaron como '1' a '6'. Los parches '1' y '6' se localizaron en el fragmento Fab, y los parches '2' a '5' se localizaron en el fragmento Fc. Los parches '1' a '3' estuvieron abiertamente expuestos y, por tanto, podrían interaccionar fácilmente con otras proteínas. Por otra parte, los parches '4' a '6' estuvieron accesibles al disolvente, pero orientados a la proteína, dificultándoles interaccionar con otras proteínas, a menos que estuvieran más abiertamente expuestos debido a desplegamiento.

A continuación se probó la correlación de regiones de alta SAP que representan parches hidrófobos expuestos con regiones de unión a proteína. Las regiones de unión del anticuerpo con receptor de Fc, proteína A y proteína G se mapearon en la parte superior de los valores de SAP. Los sitios de unión a proteína se obtuvieron de estructuras de rayos X de complejos de proteínas, entradas de PDB 1T89, 1FC2 y 1FCC (S. Radaev y col., J. Biol. Chem, 276 (19) 16469 (2001); Deisenhofer, J. y col. Hoppe-Seyler's Z. Physiol. Chem. 359, 975-985 (1978); Deisenhofer, J. Biochemistry 20, 2361-2370 (1981); Sauer-Eriksson A. E. y col., Structure, 3, 265 (1995)). Se encontró una fuerte correlación entre parches hidrófobos identificados por SAP y regiones de unión a proteína. El antígeno unido con la región de bucle de CDR marcó el parche de SAP '1', el receptor de Fc se une con el parche de SAP '2' y la proteína A y la proteína G se unen con el parche de SAP '3'. Además, DeLano y col. (DeLano W. L. y col., Science 287, 1279 (2000)) mostraron que la región en la que la proteína A y la proteína G se unen (parche de SAP '3') es una región de unión consenso que es dominante para unir péptidos aleatorios seleccionados *in vitro* para alta afinidad. También se sabe que el parche '3' se une con factor reumatoide y receptor de Fc neonatal. Por tanto, la accesibilidad hidrófoba del parche '3' como se indica por SAP hace que una región favorable se una con numerosas proteínas. Bastante sorprendentemente, los 3 parches abiertamente expuestos (parche de SAP '1' a '3') participaron en la unión. El núcleo del parche participa en interacciones hidrófobas, mientras que los flecos participan en interacciones polares.

Se analizó SAP a $R=10 \text{ \AA}$ para encontrar los anchos parches hidrófobos implicados en la unión con otras proteínas. Estos parches pueden explorarse en más detalle usando la SAP a mayor resolución, es decir, a un menor radio de R usado en el de cálculo de SAP. Por tanto, los valores de SAP se calcularon a $R=5 \text{ \AA}$ para el anticuerpo. Estos valores de SAP se mapearon sobre la superficie del anticuerpo. Aquí, los valores positivos de SAP indican parches hidrófobos dinámicamente expuestos, mientras que los valores negativos de SAP indican parches hidrófilos dinámicamente expuestos. También se identificaron regiones que se unen con receptor de Fc, proteína A y proteína G. Similar a los resultados con SAP a $R=10 \text{ \AA}$, la SAP a $R=5 \text{ \AA}$ también mostró una fuerte correlación entre regiones de unión a proteína y picos en valores de SAP. Las regiones de unión hidrófobas correlacionaron bien con los picos positivos y las regiones de unión hidrófilas (polares) correlacionaron bien con los picos negativos. En las pocas excepciones en las que los residuos con baja SAP (es decir, próxima a cero, tanto positiva como negativa) también interaccionaron, los presentes inventores observaron que las interacciones fueron en realidad con los átomos de la propia cadena principal del esqueleto, en lugar de con las cadenas laterales.

55 SAP predice tanto regiones de unión como regiones propensas a la agregación

Se ha demostrado que los picos en SAP también se corresponden con regiones que tienen tendencia a la auto-agregación de proteínas (Chennamsetty, N. y col. Design of therapeutic antibodies with enhanced stability (presentado)). La agregación es una ruta de degradación importante para proteínas terapéuticas que conduce a su pérdida de actividad y posible inmunogenicidad. Las mutaciones manipuladas sobre los picos de SAP condujeron a anticuerpos estables con menos propensión a la agregación (Chennamsetty, N. y col. Design of therapeutic antibodies with enhanced stability (presentado)). Los 8 mutantes generados cambiando los residuos hidrófobos en picos de SAP con residuos hidrófilos fueron A1 (L235K), A2 (I253K), A3 (L309K), A4 (L235K L309K), A5 (L234K L235K), A6 (L235S), A7 (V282K) y A8 (L235K V282K L309K). Los mutantes se probaron entonces para su comportamiento de agregación usando experimentos de agregación acelerados bajo tensión por calor a 150 mg/ml.

Los resultados de SEC-HPLC (cromatografía de líquidos de alta resolución con exclusión por tamaño) mostraron aumento de monómeros del 91% para natural al 92-97% para las variantes, que indica menos propensión a la agregación de los mutantes. Por tanto, los sitios con alta SAP también representan las regiones de alta propensión a la agregación.

La herramienta de SAP predijo así tanto regiones de unión a proteína como regiones propensas a la agregación. Una explicación probable es que la agregación de proteínas también es una forma de unión proteína-proteína, aunque dentro de proteínas del mismo tipo. Además, se mostró que hay un solapamiento entre algunas de las regiones propensas a la agregación y regiones de unión a proteína. Este solapamiento fue evidente de los residuos L235 y 1253 que participan en tanto la unión como la agregación de proteínas. Se realizaron análisis de SAP similares y manipulación de proteínas sobre otro anticuerpo IgG1 en el que se mostró que las regiones propensas a la agregación se solapan con regiones de unión a proteína (Chennamsetty, N. y col. Design of therapeutic antibodies with enhanced stability (presentado)). En este caso, las mutaciones se llevaron a cabo en las regiones CDR en las que el anticuerpo se une a un antígeno. Los mutantes resultantes en las regiones CDR mostraron menos propensión a la agregación, pero no pudieron unirse a un antígeno y perdieron su actividad. Así, hay características comunes con la unión a proteínas y regiones propensas a la agregación. Esto está de acuerdo con otras predicciones computacionales hechas a partir de secuencias que solapan las regiones de unión a proteína y propensas a la agregación (Wang, X. y col., mAbs, 1, 1-14 (2009)). Así, los parches hidrófobos dinámicamente expuestos identificados mediante SAP participan en tanto la unión a proteínas como la auto-agregación de proteínas.

Sin embargo, el solapamiento entre sitios de unión de proteínas y sitios propensos a la agregación presenta un nuevo reto en el diseño de proteínas terapéuticas debido a que la agregación necesita prevenirse a la vez que se preserva la unión de proteínas necesaria para su función. Para resolver este reto, el análisis de SAP a mayor resolución (a $R=5$ Å) puede usarse para localizar y modificar sitios propensos a la agregación alrededor de las regiones de unión sin alterar la unión de proteínas. Por ejemplo, usando análisis de SAP en el anticuerpo IgG1 se determinó que los sitios I253, L309 y V282 son todos parte de un amplio parche (región de SAP '3') que participan en la agregación (Chennamsetty, N. y col. Design of therapeutic antibodies with enhanced stability (presentado)). Se diseñaron mutantes que implican sitios L309 y V282 {A3 (L309K), A4 (L235K L309K), A7 (V282K) y A8 (L235K V282K L309K)}, excluyendo el sitio I253 que se implicó en la unión a proteína A. Los mutantes resultantes mostraron menos propensión a la agregación, aunque todavía se unen a proteína A. Así, la tecnología de SAP puede usarse eficazmente para diseñar proteínas con una menor propensión a la agregación mientras que se preserva la capacidad de unión a proteína.

SAP predice regiones de unión de EGFR

Además de anticuerpos, el análisis de SAP se realizó sobre otra proteína llamada receptor de factor de crecimiento epidérmico (EGFR) para predecir sus regiones de unión. El EGFR es un receptor de la superficie celular activado por la unión de ligandos específicos que incluyen receptor de factor de crecimiento epidérmico (EGF) y factor de crecimiento transformante β (TGF β). La expresión en exceso o actividad en exceso de EGFR se ha asociado a varios cánceres tales como cáncer de pulmón y cáncer cerebral. El EGFR también se une consigo mismo para formar dímeros. Se realizó un análisis de SAP sobre EGFR para ver si las regiones de unión predichas coincidían con las regiones de unión de EGF, TGF α , y con otra EGFR en la forma dímera.

Los valores de SAP evaluados para EGFR a $R=10$ Å se mapearon sobre la superficie de EGFR. Estos valores de SAP se calcularon realizando el análisis directamente sobre estructura de rayos X de EGFR obtenida de la entrada de PDB 1IVO (Ogiso, H. y col., Cell, 110: 775-787 (2002)). La escala de hidrofobia (S. D. Black y D. R. Mould, Anal. Biochem. 193, 72 (1991)) también se mapeó sobre la superficie de EGFR para comparación. Como se ha observado anteriormente en el caso del anticuerpo, los residuos hidrófobos para EGFR se distribuyeron por toda la superficie, y sería difícil aislar los potencialmente implicados en la unión. Sin embargo, fue relativamente más fácil localizar las regiones de alta SAP, que indican regiones hidrófobas espacialmente expuestas. Dos parches principales tales se identificaron y se marcaron '1' y '2'.

Las regiones de unión conocidas de EGFR con EGF, TGF α y con otro EGFR en la forma dímera se mapearon en la parte superior de los valores de SAP. Estos sitios de unión a proteína se obtuvieron de estructuras de rayos X de complejos de proteína, entradas de PDB 1IVO y 1MOX (Ogiso, H. y col. Cell, 110: 775-787 (2002); Garrett, T.P.J. y col. Cell, 110: 763-773 (2002)). El mapeo indicó una fuerte correlación entre parches hidrófobos identificados por SAP y regiones de unión a proteína. EGFR se une con EGF y TGF α en el parche de SAP '1' y otro parche más pequeño. También se une con otro EGFR en el parche de SAP '2'. Así, los dos parches de SAP principales están ambos implicados en la unión. De nuevo como en el caso del anticuerpo, el núcleo de los parches participa en interacciones hidrófobas, mientras que los flecos participan en interacciones polares. Así, la SAP predijo con exactitud las regiones de unión de EGFR.

Conclusiones

Se ha descrito una herramienta computacional llamada SAP que proporciona una medida de la exposición dinámica de parches hidrófobos que puede usarse para predecir regiones de unión a proteína. Usando dos proteínas modelo,

un anticuerpo IgG1 y EGFR, se mostró que SAP predice con exactitud regiones de unión a proteína. En el caso del anticuerpo IgG1, las regiones de unión con receptor de Fc, proteína A y proteína G se correlacionaron bien con picos de SAP. Para EGFR, las regiones de unión con EGF, TGFβ y con otro EGFR correlacionaron bien con picos de SAP. Así, se mostró que SAP era exacto en predecir regiones de unión, y se demostró la importancia de parches hidrófobamente expuestos para la unión proteína-proteína. El mismo análisis de SAP podría realizarse sobre otras proteínas para también predecir sus regiones de unión. Además, se ha mostrado que algunas de las regiones de unión a proteína se solapan con regiones propensas a la agregación. Esto presenta un reto para el diseño de proteínas terapéuticas debido a que debe evitarse la agregación desfavorable mientras que se preserve la unión a proteína necesaria para su función. Se ha mostrado que este reto puede superarse usando análisis de SAP, seguido de manipulación de proteínas. Usando SAP, los sitios próximos al sitio de unión que participan en la agregación pueden detectarse y modificarse para reducir la propensión a la agregación mientras que se preserva la unión. Esto se demostró usando el anticuerpo IgG1 en el que las regiones propensas a la agregación próximas a los sitios de unión a la proteína A se modificaron para reducir la agregación, mientras que se preservaba la capacidad de unión. La manipulación de proteínas similares basada en SAP podría realizarse próxima a las regiones de unión a antígeno para reducir la propensión a la agregación, mientras que se preserva la actividad. Así, la herramienta de SAP descrita aquí podría usarse para diseñar proteínas terapéuticas estables, mientras que al mismo tiempo se preservan sus sitios de unión. La herramienta de SAP también podría usarse para determinar los sitios de unión todavía desconocidos para numerosas proteínas procedentes de iniciativas de genómica estructural, proporcionándose así importantes pistas para su función.

Aquellos expertos en la materia reconocerán, o podrán determinar usando no más de experimentación rutinaria, muchos equivalentes a las realizaciones específicas de la invención descritas en el presente documento. Tales equivalentes pretenden estar englobados por las siguientes reivindicaciones.

25 LISTADO DE SECUENCIAS

<110> Novartis AG Massachusetts Institute of Technology CHENNAMSETTY, NARESH HELK, BERNHARD TROUT, BERNHARDT L.

30 <120> PROCEDIMIENTOS PARA IDENTIFICAR REGIONES DE UNIÓN A MACROMOLÉCULA Y PROPENSAS A LA AGREGACIÓN EN PROTEÍNAS Y USOS DE LOS MISMOS

<130> 61967-2000340

35 <140> Todavía sin asignar
<141> Simultáneamente con la presente

<150> US 61/074.466
<151> 20/06/2008

40 <160> 3

<170> FastSEQ para Windows Versión 4.0

45 <210> 1
<211> 7
<212> PRT
<213> Saccharomyces cerevisiae

50 <400> 1

Gly Asn Asn Gln Gln Asn Tyr
1 5

55 <210> 2
<211> 7
<212> PRT
<213> Homo sapiens

60 <400> 2

Lys Leu Val Phe Phe Ala Glu
1 5

<210> 3
<211> 11
<212> PRT
65 <213> Homo sapiens

ES 2 417 781 T3

<400> 3

Val His His Gln Lys Leu Val I
1 5

5

REIVINDICACIONES

- 5 1. Un procedimiento implementado por ordenador para identificar una región propensa a la agregación y/o una
región de unión a macromolécula sobre una proteína que comprende (a) mapear, sobre un modelo estructural de la
proteína, la propensión a la agregación espacial, SAP, para átomos en la proteína como se calcula para un átomo
particular
- 10 (i) identificando uno o más átomos o residuos de aminoácidos en un modelo estructural que representa la
proteína, en el que el uno o más átomos están dentro de una región espacial definida centrada en o
próxima al átomo particular o el uno o más residuos de aminoácidos tienen uno o más átomos dentro de
una región espacial definida centrada en o próxima al átomo particular;
- 15 (ii) calculando, para el uno o más átomos en la región espacial definida, una relación del área accesible al
disolvente, SAA, de los átomos con respecto al SAA de átomos en un residuo idéntico que está
completamente expuesto;
- (iii) multiplicando cada relación por la hidrofobia del átomo del uno o más átomos o la hidrofobia del uno o
más residuos de aminoácidos como se ha determinado por una escala de hidrofobia de aminoácidos; y
(iv) sumando los productos de la etapa (iii); por lo que la suma es la SAP para el átomo particular; o
- 20 en el que la SAP para el átomo particular se calcula realizando una simulación de dinámica molecular antes de la
etapa (i) y repitiendo las etapas (i)-(iv), realizando cada vez otra simulación de dinámica molecular en una pluralidad
de etapas de tiempo, produciendo así múltiples sumas como en la etapa (iv), y calculando el promedio de las sumas;
por lo que el promedio calculado es la SAP para el átomo particular; y (b) identificar una región dentro de la proteína
que tiene una pluralidad de átomos que tiene una SAP > 0;
- 25 en el que la región propensa a la agregación y/o región de unión a macromolécula comprende los aminoácidos que
comprenden dicha pluralidad de átomos.
2. El procedimiento implementado por ordenador de la reivindicación 1, en el que la identificación comprende
representar los valores de SAP; calcular, para picos en la representación, el área bajo la curva, ABC; e identificar
30 una o más regiones de proteína con un ABC positiva, en el que la región propensa a la agregación y/o región de
unión a macromolécula comprende las regiones de proteína identificadas.
3. El procedimiento implementado por ordenador de la reivindicación 1 o la reivindicación 2 para su uso en un
procedimiento de preparación de una variante de proteína que presenta una propensión reducida a la agregación y/o
35 afinidad de unión alterada por una macromolécula, en el que el procedimiento de preparación de una variante de
proteína comprende sustituir o deleccionar al menos un residuo de aminoácido dentro de una región propensa a la
agregación en la proteína,
en el que la región propensa a la agregación se identifica usando puntuaciones de SAP calculadas según la
reivindicación 1; y
- 40 en el que, si el residuo de aminoácido se sustituye, se sustituye con un residuo de aminoácido que es más hidrófilo,
de forma que se reduce la propensión a la agregación de la variante y/o la afinidad de unión por la macromolécula
de la variante.
4. El procedimiento implementado por ordenador de la reivindicación 1 o la reivindicación 2 para su uso en un
procedimiento de preparación de una variante de proteína que presenta una propensión reducida a la agregación y/o
45 afinidad de unión alterada por una macromolécula, en el que el procedimiento de preparación de una variante de
proteína comprende
- 50 (a) generar una pluralidad de variantes de proteína sustituyendo en cada variante al menos un residuo
dentro de una región propensa a la agregación en la proteína, en el que la región propensa a la agregación
se identifica usando puntuaciones de SAP calculadas según la reivindicación 1,
en el que un residuo o residuos diferentes, o combinaciones diferentes de residuos, se sustituyen en cada
variante;
- 55 en el que el al menos un residuo se sustituye con un residuo que es más hidrófilo; y
(b) seleccionar una variante de proteína preparada como en (a) que presenta una propensión reducida a la
agregación y/o que presenta una afinidad de unión alterada por la macromolécula.
5. El procedimiento implementado por ordenador de la reivindicación 1 o la reivindicación 2 para su uso según la
reivindicación 3 o la reivindicación 4, en el que se sustituyen al menos dos residuos de aminoácidos dentro de la
60 región propensa a la agregación y/o región de unión a macromolécula.
6. El procedimiento implementado por ordenador de la reivindicación 1 o la reivindicación 2 para su uso según la
reivindicación 3 o la reivindicación 4, en el que al menos un residuo se sustituye dentro de más de una región
propensa a la agregación y/o más de una región de unión a macromolécula dentro de la proteína.
- 65 7. El procedimiento de una cualquiera de la reivindicación 1 a la reivindicación 6, en el que la proteína está

seleccionada del grupo constituido por un anticuerpo, un fragmento Fab, un fragmento Fab', un fragmento Fd, un fragmento Fv, un fragmento F(ab')₂ y un fragmento Fc.

- 5 8. Un procedimiento para la preparación de una composición farmacéutica que comprende una variante de proteína que presenta una propensión reducida a la agregación y/o una propensión reducida para la interacción con un componente de unión, que comprende formular una variante de proteína obtenida según el procedimiento de la reivindicación 3 o la reivindicación 4 junto con un vehículo, adyuvante y/o excipiente farmacéuticamente aceptable.
- 10 9. Un medio legible por ordenador que comprende instrucciones, que cuando se ejecuta hace que un procesador lleve a cabo el procedimiento implementado por ordenador según la reivindicación 1 o la reivindicación 2.