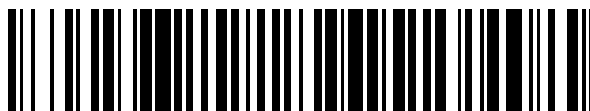


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 701 440**

51 Int. Cl.:

G06F 19/16

(2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **16.08.2013** **PCT/CA2013/050637**

87 Fecha y número de publicación internacional: **20.02.2014** **WO14026296**

96 Fecha de presentación y número de la solicitud europea: **16.08.2013** **E 13829560 (5)**

97 Fecha y número de publicación de la concesión europea: **03.10.2018** **EP 2885736**

54 Título: **Sistemas y métodos de muestreo y análisis de dinámica conformacional de polímeros**

30 Prioridad:

17.08.2012 US 201261684236 P

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

22.02.2019

73 Titular/es:

**ZYMEWORKS, INC. (100.0%)
540-1385 West 8th Avenue
Vancouver, British Columbia V6H 3V9, CA**

72 Inventor/es:

DIXIT, SURJIT B.

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 701 440 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

Sistemas y métodos de muestreo y análisis de dinámica conformacional de polímeros

5 **Campo técnico**

La invención proporciona sistemas y métodos para buscar el espacio de conformación de un polímero para determinar una conformación tridimensional del polímero que cumple una métrica de rendimiento.

10 **Antecedentes**

Los polímeros son moléculas altamente dinámicas y muchas de sus características funcionalmente importantes, tales como la afinidad por sustratos y la estabilidad de la forma activa, dependen de un conjunto de las estructuras que comprenden múltiples subestados conformacionales, sus probabilidades y tasas de transición, y características de la superficie libre de energía intrínseca.

En el caso de las proteínas, hay una creciente evidencia que indica que la unión de una proteína a objetivos particulares a menudo se asocia a una selección preferente de una o varias de estas conformaciones. Ver Smock y Gierasch, 2009, "Sending signals dynamically", *Science* 324 (5924): 198-203.

La información estructural detallada obtenida mediante procedimientos experimentales, tales como cristalografía de rayos X, no proporciona mucha información sobre el comportamiento dinámico de un polímero. La comprensión de la dinámica de los polímeros es especialmente importante en los estudios de los efectos ambientales y de los solventes sobre la estabilidad y la asociación macromolecular, procesos que suelen ir acompañados de una reorganización estructural.

Con el fin de desarrollar una comprensión más amplia de cómo funcionan los polímeros, y para desarrollar mejores polímeros industriales y farmacéuticos, es necesario desarrollar un aprecio por estos diversos estados conformacionales. Tradicionalmente, el muestreo conformacional de estructuras de polímeros se lleva a cabo utilizando enfoques deterministas o de simulación estocástica. Ver Adcock McCammon, 2006, "Molecular dynamics: survey of methods for simulating the activity of proteins" *Chem Rev* 106 (5):1589-615. Una divulgación a modo de ejemplo, el documento WO 2002/079784 A1, divulga un sistema y un método para buscar un espacio combinatorio mediante la búsqueda de coordenadas dentro de tal espacio, y luego alterar un polímero pivotando las hélices individuales dentro de un dominio de polímero individual. Sin embargo, esta enseñanza solo considera las moléculas en las que se produce el pivote alrededor de un único átomo.

En un enfoque determinista, tal como simulación de dinámica molecular, La mecánica newtoniana se emplea para calcular la trayectoria de todas las partículas en el sistema como resultado de las fuerzas de interacción que actúan entre ellas. En este procedimiento, los desplazamientos atómicos se estiman en etapas de tiempo muy cortas, y la integración numérica se lleva a cabo en un cálculo iterativo para predecir la dinámica detallada del sistema en escalas de tiempo más largas. Hay un límite superior para la etapa de tiempo que se puede usar, y esto, junto con el gran tamaño del sistema de polímeros en estudio, limita computacionalmente la longitud de la simulación que se puede realizar. Esto, a su vez, limita el número de transiciones conformacionales a gran escala y, por lo tanto, el número de subestados del polímero que pueden revelarse en el curso de una única trayectoria. Aunque estas simulaciones computacionales tradicionales abordan el carácter dinámico de los polímeros, los enfoques de dinámica molecular (deterministas) son computacionalmente intensos, haciéndolos poco prácticos para estudiar cambios conformacionales lentos de polímeros más grandes, tales como proteínas grandes.

En un enfoque estocástico, tal como el muestreo de Monte Carlo, una serie de variables en el sistema se seleccionan aleatoriamente y se perturban para generar una nueva configuración del sistema. En una versión evolucionada del algoritmo, tal como el método de Metropolis Monte Carlo, la nueva configuración se acepta o se rechaza basándose en un criterio energético a la temperatura de interés, lo que lleva a un conjunto ponderado de Boltzmann de configuraciones termodinámicamente relevantes. El uso del muestreo de Monte Carlo puede dar como resultado saltos significativamente más eficaces entre los estados conformacionales relevantes, superando así las barreras observadas en las simulaciones de dinámica molecular tradicional. Pero seleccionando los grados de libertad que, cuando se perturba, daría lugar a una probabilidad de aceptación factible en el marco del esquema Metropolis a menudo es un problema importante en la aplicación de este enfoque. En otras palabras, las simulaciones basadas en algoritmos puramente estocásticos, tales como la técnica de Metropolis Monte-Carlo, no han logrado resolver este problema porque tienden a producir relaciones de aceptación limitadas que dan como resultado una simulación ineficiente. Los métodos que incluyen movimientos estocásticos con la evolución determinista MD se han tratado, pero se han limitado en su aplicación. Ver, Guamieri y Still, 1994, "A rapidly convergent simulation method: Mixed Monte Carlo/stochastic dynamics", *J Comput Chem* 15 (11):1302-1310; y la patente estadounidense № 5,740,072. Otra publicación (Flores y Gerstein, 2011 "Predicting protein ligand binding motions with the conformation explorer", *BMC Bioinformatics* 12, pág. 417), describe métodos para evaluar las estructuras tridimensionales candidatas de los polipéptidos por dinámica molecular. Sin embargo, esta enseñanza no considera la puntuación de una energía potencial calculada para cada estructura alterada frente a la no rotada inicial para determinar si la estructura polimérica

alterada cumple con un criterio de Metrópolis deseado particular.

Teniendo en cuenta el contexto anterior, se necesitan, por lo tanto, sistemas y un método mejorados para llegar a la información conformacional del polímero.

5

Sumario de la presente divulgación

Se ha desarrollado un enfoque a la cuestión del muestreo conformacional de un polímero usando la percepción de un modelo de alto nivel o de grano grueso del polímero y el análisis heurístico de la estructura del polímero, para inferir la ubicación de los dominios, bucles, y regiones de bisagra. Basándose en este análisis, se prevén cambios conformacionales específicos a gran escala y los estados conformacionales asociados del polímero. Luego, se realizan simulaciones atomísticas detalladas para desarrollar una comprensión refinada de los estados conformacionales. Esto revela el panorama energético y conformacional del polímero.

10

La presente invención se realiza en las reivindicaciones adjuntas. Por tanto, en un aspecto, se proporcionan métodos para muestrear estados conformacionales de un polímero usando un sistema informático en el que se realiza un modelado de grano grueso del polímero para generar una conformación alterada y, luego, se realiza una simulación atomística de la conformación alterada. En algunas realizaciones, el modelado de grano grueso comprende la realización de un método computacional estocástico, tal como, por ejemplo, Monte Carlo. En algunas realizaciones, la simulación atomística comprende realizar un método computacional determinista, tal como, por ejemplo, dinámica molecular.

15

20

Los sistemas y métodos descritos en el presente documento proporcionan una rápida generación y análisis de múltiples estados conformacionales de polímeros potencialmente importantes. Los métodos de búsqueda deterministas y estocásticos se combinan de una manera que ofrece un muestreo conformacional mejorado de polímeros en comparación con el uso de cualquiera de estas dos técnicas por sí mismas. Los sistemas y métodos divulgados incorporan la fuerza de los modelos de grano grueso y el conocimiento heurístico de polímeros para predecir las ubicaciones de las bisagras en un polímero. Monte Carlo se mueve para introducir cambios conformacionales a gran escala y los combina con la fuerza de una simulación de Dinámica Molecular para proporcionar información detallada sobre la flexibilidad conformacional alrededor de un sustrato estructural. Los sistemas y métodos divulgados permiten encontrar diversas conformaciones de polímeros y sus poblaciones relativas al mismo tiempo, llevando a una sorprendente velocidad y practicidad de modelado a la luz de métodos y soluciones anteriores.

25

30

Un aspecto de la presente divulgación proporciona un método para buscar el espacio de conformación de un polímero para determinar una conformación tridimensional del polímero que satisface un criterio de Metropolis. El método comprende, en un sistema informático que tiene uno o varios procesadores y memoria que almacena uno o varios programas a ejecutar por el uno o varios procesadores, obtener un conjunto inicial de coordenadas tridimensionales $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$ para el polímero. El polímero comprende una pluralidad de dominios. Cada respectivo $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$ es una coordenada tridimensional para un átomo en un primer dominio en la pluralidad de dominios. Cada respectivo X_{iB} en $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$ es una coordenada tridimensional para un átomo en un segundo dominio en la pluralidad de dominios. Cada respectivo X_{iC} en $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$ es una coordenada tridimensional para un átomo en una primera bisagra del polímero. El polímero está caracterizado por una capacidad para que el primer y el segundo dominio pivoten entre sí alrededor de la primera bisagra. El conjunto de coordenadas tridimensionales del polímero se altera pivotando el primer dominio con respecto al segundo dominio sobre la primera bisagra, obteniendo así un conjunto alterado de coordenadas tridimensionales $\{X_{1A_alt}, \dots, X_{NA_alt}, X_{1B_alt}, \dots, X_{MB_alt}, \dots, X_{PC_alt}, \dots\}$ para el polímero. Durante esta etapa de alteración, los átomos dentro del primer dominio se mantienen fijos entre sí y los átomos dentro del segundo dominio se mantienen fijos entre sí. El conjunto alterado de coordenadas se puntúa según el criterio de Metropolis. Se realizan instancias adicionales de la alteración y de la puntuación hasta que el conjunto alterado de coordenadas tridimensionales $\{X_{1A_alt}, \dots, X_{NA_alt}, X_{1B_alt}, \dots, X_{MB_alt}, \dots, X_{PC_alt}, \dots\}$ satisfaga el criterio de Metropolis.

35

40

45

50

En algunas realizaciones, antes de la etapa inicial de alteración, se determina qué residuos del polímero se encuentran en el primer dominio, la primera bisagra, y el segundo dominio.

55

En algunas realizaciones, el polímero comprende una pluralidad de bisagras. Esta pluralidad de bisagras incluye la primera bisagra. Para cada bisagra respectiva en la pluralidad de bisagras, hay un par correspondiente de dominios en la pluralidad de dominios que pivotan entre sí en torno a la bisagra respectiva. En algunas tales realizaciones, el método comprende, además, antes de la etapa de alteración, determinar la identidad de los residuos en cada bisagra en la pluralidad de bisagras. En algunas realizaciones, esta información ya es conocida. En algunas realizaciones, la alteración hace pivotar, además, un tercer dominio con respecto a un cuarto dominio en la pluralidad de dominios alrededor de una segunda bisagra en la pluralidad de bisagras. Al hacerlo, los átomos dentro del tercer dominio se mantienen fijos entre sí y los átomos dentro del cuarto dominio se mantienen fijos entre sí.

60

Se divulga que la puntuación comprende resolver un problema de cierre de bucle para que la primera bisagra introduzca simultáneamente alteraciones en algunos o en la totalidad de una pluralidad de parámetros de bisagra

65

asociados a la primera bisagra. Se divulga que menos de $2(n-2)$ parámetros de bisagra se alteran en el problema de cierre del bucle. En algunas realizaciones, n es el número de residuos en la primera articulación.

En algunas realizaciones, la ubicación de la primera bisagra se determina sometiendo el conjunto inicial de coordenadas tridimensionales al análisis de modo normal, análisis de componentes principales, análisis de covarianza, análisis de dominio de proteína, análisis de rigidez, análisis de características de secuencia.

En algunas realizaciones, determinar la identidad de los residuos en cada bisagra en la pluralidad de bisagras comprende someter el conjunto inicial de coordenadas tridimensionales al análisis de modo normal, análisis de componentes principales, análisis de covarianza, análisis de dominio de proteína, análisis de rigidez, análisis de características de secuencia.

En algunas realizaciones, el polímero es una proteína y la alteración comprende, además, el reempaquetamiento de geometrías de cadena lateral de proteínas en el conjunto alternativo de coordenadas tridimensionales aplicando un algoritmo de empaquetamiento de proteínas. En algunas realizaciones, el algoritmo de empaquetamiento de proteínas optimiza la geometría de cadena lateral para los aminoácidos en la proteína que son desplazados por la etapa de alteración.

Se divulga que la métrica de rendimiento comprende una función de energía basada en la física. Se divulga que la métrica de rendimiento comprende una función de energía basada en el conocimiento.

En algunas realizaciones, la alteración y la puntuación se llevan a cabo secuencialmente en el tiempo y en donde una enésima instancia de la alteración altera el conjunto alterado de coordenadas del polímero desde la instancia enésima-1 o enésima-2 de la alteración. Esta selección del conjunto alterado de coordenadas del polímero para la enésima instancia de la alteración se basa en los criterios de Metropolis, y n es un número entero positivo de tres o mayor.

El polímero es una proteína, la primera bisagra comprende una pluralidad de residuos, cada residuo en la pluralidad de residuos está asociado a un ángulo de torsión de la red troncal en una pluralidad de ángulos de torsión de la red troncal, y el giro del primer dominio con respecto al segundo dominio comprende polarizar los respectivos ángulos de torsión de la red troncal en la pluralidad de ángulos de torsión de la red troncal hacia estados diedros estereoquímicamente aceptables.

En algunas realizaciones, el método comprende, además, llevar a cabo una simulación de dinámica molecular del conjunto alterado de coordenadas tridimensionales antes de la puntuación. En algunas realizaciones, esta simulación de dinámica molecular comprende el uso de un modelo de solvente explícito o implícito. En algunas realizaciones, esta simulación de dinámica molecular es una simulación de dinámica molecular de intercambio de réplica en serie (SRE). En algunas realizaciones, la simulación de dinámica molecular SRE comprende templado de temperatura o templado de solvente.

En algunas realizaciones, la puntuación comprende calcular un panorama energético libre basado en el conjunto alterado de coordenadas tridimensionales utilizando un método seleccionado del grupo que consiste en Muestreo de Paraguas, Integración Termodinámica, Perturbación de Energía Libre, Fuerza de Polarización Adaptativa basada en el potencial de los cálculos de fuerza media y dinámica molecular dirigida.

En algunas realizaciones, se lleva a cabo una pluralidad de instancias de alteración generando así una pluralidad de conjuntos alterados de coordenadas tridimensionales.

El polímero es una proteína y el estado de protonación de una cadena lateral de aminoácido ionizable en el polímero varía entre la pluralidad de conjuntos alterados de coordenadas tridimensionales.

En algunas realizaciones, el método comprende, además, la asignación de cada conjunto alterado de coordenadas tridimensionales en la pluralidad de conjuntos alterados de coordenadas tridimensionales a uno de una pluralidad de clústeres conformacionales. La etapa de asignación comprende la agrupación basada en datos estructurales de la pluralidad de conjuntos alterados de coordenadas tridimensionales seleccionadas del grupo que consiste en un parámetro geométrico interdominio, un patrón de contacto interatómico, un radio de rotación, una accesibilidad de solvente y la desviación cuadrática media.

En algunas realizaciones, una pluralidad de instancias de la alteración se realiza de manera simultánea, obteniendo así una pluralidad de conjuntos alterados de coordenadas tridimensionales para el polímero. En algunas tales realizaciones, el método comprende, además, realizar simultáneamente una simulación independiente de dinámica molecular de cada conjunto alterado de coordenadas tridimensionales en la pluralidad de conjuntos alterados de coordenadas tridimensionales. De este modo, se realiza una pluralidad de simulaciones de dinámica molecular independientes, antes de la etapa de puntuación. En algunas realizaciones, la pluralidad de simulaciones de dinámica molecular independientes se realiza en paralelo (a) en nodos de un clúster de informático, (b) en un sistema informático distribuido o (c) mediante computación de propósito general en unidades de procesamiento de gráficos. En algunas realizaciones, cuando una similitud conformacional entre estructuras intermedias en cualquiera de los dos de la

pluralidad de simulaciones de dinámica molecular independiente está dentro de un umbral predeterminado, una de las dos simulaciones de dinámica molecular está terminada. En algunas realizaciones, una simulación de dinámica molecular independiente en la pluralidad de simulación de dinámica molecular independiente comprende el cálculo adaptativo de una distribución de energía de una pluralidad de conformaciones del polímero.

5 En algunas realizaciones, la alteración sesga el conjunto alterado de coordenadas tridimensionales hacia un modo de baja frecuencia principal de movimiento mediante el análisis de componentes principales de una trayectoria de dinámica molecular o un análisis de modelo de red elástica anisotrópica.

10 En algunas realizaciones, el polímero se caracteriza por una secuencia primaria de residuos y los residuos en la primera bisagra son de diferentes porciones de la secuencia primaria, que se interrumpen por porciones de la secuencia primaria que están en el primer dominio o en el segundo dominio.

15 El polímero puede ser un polipéptido. En algunas realizaciones, el polímero comprende treinta o más residuos, cuarenta o más residuos, cincuenta o más residuos, o uno o varios residuos. En algunas realizaciones, la primera bisagra comprende cinco o más residuos, diez o más residuos, quince o más residuos, o veinte o más residuos.

20 El polímero es una proteína, la primera bisagra comprende una pluralidad de residuos, cada residuo en la pluralidad de residuos está asociado a un ángulo de torsión de la red troncal en una pluralidad de ángulos de torsión de la red troncal, el giro del primer dominio con respecto al segundo dominio comprende polarizar los respectivos ángulos de torsión de la red troncal en la pluralidad de ángulos de torsión de la red troncal hacia estados diedros estereoquímicamente aceptables, y el polímero se caracteriza por una secuencia primaria de residuos. Los residuos en la primera bisagra son de diferentes porciones de la secuencia primaria, que se interrumpen por porciones de la secuencia primaria que están en el primer dominio o en el segundo dominio.

25 En algunas realizaciones, la alteración y la puntuación se realizan utilizando un enfoque de Monte Carlo.

30 El polímero es una proteína, la primera bisagra comprende n residuos, donde n es un número entero positivo de 2 o mayor, la primera bisagra comprende $2(n-2)$ ángulos diedros de la red troncal, un subconjunto de $2(n-2)$ ángulos diedros de red troncal en la primera bisagra no se altera durante la etapa de alteración, y la etapa de alteración y la etapa de puntuación se realizan utilizando un enfoque de Monte Carlo.

35 En algunas realizaciones, el método comprende, además, refinar el conjunto alterado de coordenadas tridimensionales $\{X1A_alt, \dots, XNA_alt, X1B_alt, \dots, XMB_alt, \dots, XPC_alt, \dots\}$ utilizando la dinámica molecular.

En algunas realizaciones, la puntuación se lleva a cabo utilizando una función de energía potencial. En algunas realizaciones, la puntuación se realiza utilizando una función de energía basada en el conocimiento.

40 Otro aspecto de la presente divulgación proporciona un sistema informático para buscar el espacio de conformación de un polímero para determinar una conformación tridimensional del polímero que satisface un criterio de Metropolis, comprendiendo el sistema informático al menos un procesador y una memoria que almacena al menos un programa para ser ejecutado por al menos un procesador, la memoria comprende, además, instrucciones para ejecutar el método de cualquiera de los métodos descritos en el presente documento.

45 Otro aspecto de la presente divulgación proporciona un medio de almacenamiento legible por ordenador no transitorio que almacena uno o varios módulos computacionales para buscar el espacio de conformación de un polímero para determinar una conformación tridimensional del polímero que satisface un criterio de Metropolis, comprendiendo el uno o varios módulos computacionales colectivamente instrucciones para llevar a cabo cualquiera de los métodos descritos en el presente documento.

50 Breve descripción de los dibujos

55 Las realizaciones descritas en el presente documento se ilustran a modo de ejemplo, y no a modo de limitación, en las figuras de los dibujos adjuntos. Los números de referencia similares se refieren a las partes correspondientes en todos los dibujos.

La figura 1 es un diagrama de bloques que ilustra un sistema, de acuerdo con una forma de realización a modo de ejemplo de la presente divulgación.

60 La figura 2 ilustra un método para buscar el espacio de conformación de un polímero para determinar una conformación tridimensional del polímero que satisface un criterio de Metropolis de acuerdo con una realización de la presente divulgación.

65 Las Figuras 3 y 4 muestran una representación gráfica de la información direccional para los residuos en una proteína con respecto a los diferentes modos de movimiento derivados del análisis de modo normal. Cuando los residuos adyacentes tienen la misma dirección, su movimiento es concertado como grupo. Cuando la dirección cambia de signo entre residuos vecinos, indica una unión entre dos dominios que se mueven en diferentes direcciones y una posible bisagra. Las magnitudes de los componentes de desplazamiento se usan para calibrar

la importancia de diversas posiciones como bisagras potenciales.

La figura 5 muestra un ejemplo de información de covarianza, que puede derivarse del Modelo de Red Elástica, métodos de Análisis de Componente Principal (PCA) o Mapa de Correlación Cruzada Dinámica (DCCM). La información de covarianza proporcionada por estos métodos se puede usar para identificar posibles bisagras.

La figura 6 muestra una ilustración de las posibles posiciones de la bisagra predichas (sombreado claro) mapeadas en la secuencia de un ejemplo de proteína (Ribonucleasa: PDB ID 1BU4). Las cinco imágenes representan las bisagras predichas en los cinco modos de movimiento de frecuencia más baja para esta proteína, según lo identificado por el análisis de GNM.

La figura 7 muestra una representación esquemática de una molécula de anticuerpo IgG. Es una estructura homodimérica unida por enlaces de disulfuro e interacciones no covalentes entre diversos dominios. La bisagra conecta la región Fab a la porción Fc de la molécula de IgG. La bisagra tiene una longitud de 15 aminoácidos en el caso de IgG1, 12 aminoácidos de largo en IgG2 e IgG4 y 64 aminoácidos de largo en IgG3.

La figura 8 muestra dos scFv fusionados con un albucore que usa dos enlazadores de acuerdo con una realización a modo de ejemplo de la presente divulgación. El enlazador funciona como una bisagra entre el albucore y los dos scFv.

La figura 9 muestra una distribución de histograma de ángulos aceptados para tres categorías de residuos de bisagra: base de la bisagra, centro de la bisagra y uno más cercano al dominio en movimiento. Los ángulos phi y psi se trazan por separado. Los ángulos aceptados para las torsiones más cercanas a la base de la bisagra son significativamente más pequeños que los aceptados más cercanos a la región móvil. Esto se racionaliza sobre la base de que

se mueve en ángulos de torsión más cercanos a la base de la bisagra resultarán en rotaciones del cuerpo fijas mucho más grandes.

La figura 10 muestra la probabilidad de distancia entre B/232 y A/395 en una molécula de proteína basada en albúmina de tipo salvaje de acuerdo con un ejemplo de la presente divulgación.

La figura 11 muestra una distribución de histograma de la distancia entre B/232 y A/395 usando la misma estructura usada para la Figura 10, con la excepción de que los residuos de bisagra han mutado a prolina.

La figura 12 muestra una distribución de histograma de la distancia entre B/232 y A/395 usando la misma estructura usada para la figura 10, con la excepción de que todos los residuos distintos de los residuos de bisagra han mutado a alanina.

La figura 13 proporciona una captura de pantalla de una estructura alternativa de acuerdo con algunas realizaciones.

La figura 14 proporciona una captura de pantalla de una estructura alternativa de acuerdo con algunas realizaciones.

La figura 15 proporciona una ilustración de un complejo Fc de acuerdo con algunas realizaciones.

La figura 16 proporciona un histograma para una cadena de carbohidratos en la estructura Fc de la figura 15.

La figura 17 proporciona un histograma para otra cadena de carbohidratos en la estructura Fc de la figura 15.

La figura 18 es un ejemplo de un guion para el muestreo de bisagra zapp de acuerdo con una realización a modo de ejemplo.

La figura 19 es un guion a modo de ejemplo utilizado para evaluar un polímero de acuerdo con una realización a modo de ejemplo.

Descripción detallada

Se divulgan métodos para alterar una conformación de entrada de un polímero para generar una conformación alterada del polímero en un sistema informático. Otro aspecto proporciona un método para alterar una conformación de entrada de un polímero para generar una conformación alterada del polímero en un sistema informático, en el que el polímero comprende (i) una pluralidad de aminoácidos y (ii) una bisagra caracterizada por un parámetro de bisagra, comprendiendo el método (a) determinar la ubicación de la bisagra y (b) aplicar una alteración del parámetro de la bisagra, generando de este modo la conformación alterada del polímero.

Lo siguiente proporciona sistemas y métodos de búsqueda del espacio de conformación de un polímero para determinar una conformación tridimensional del polímero que satisface un criterio de Metropolis. La figura 1 es un diagrama de bloques que ilustra un sistema 11 de acuerdo con una de tales realizaciones. El sistema 11 comprende un ordenador 10 que normalmente incluye una o varias unidades de procesamiento (CPU, a veces llamadas procesadores) 22 para ejecutar programas (por ejemplo, programas almacenados en la memoria 36), una o varias interfaces de red u otras interfaces de comunicaciones 20, memoria 36, una interfaz de usuario 32, que incluye uno o varios dispositivos de entrada (tal como un teclado 28, un ratón 72, una pantalla táctil, teclados numéricos, etc.) y uno o varios dispositivos de salida, tales como un dispositivo de visualización 26, y uno o varios buses de comunicación 30 para interconectar estos componentes. Los buses de comunicación 30 pueden incluir circuitería (a veces llamada chipset) que interconecta y controla las comunicaciones entre los componentes del sistema.

La memoria 36 incluye memoria de acceso aleatorio de alta velocidad, tal como DRAM, SRAM, DDR RAM u otros dispositivos de memoria de estado sólido de acceso aleatorio; y comúnmente incluye memoria no volátil, tal como uno o varios dispositivos de almacenamiento de disco magnético, dispositivos de almacenamiento de disco óptico, dispositivos de memoria flash, u otros dispositivos de almacenamiento de estado sólido no volátiles. La memoria 36 opcionalmente incluye uno o varios dispositivos de almacenamiento ubicados remotamente desde la(s) CPU 22. La

memoria 36, o alternativamente, el (los) dispositivo(s) de memoria no volátil dentro de la memoria 36, comprende un medio de almacenamiento legible por ordenador no transitorio. En algunas realizaciones, la memoria 36 o el medio de almacenamiento legible por ordenador de la memoria 36 almacena los siguientes programas, módulos y estructuras de datos, o un subconjunto de los mismos:

- 5 • un sistema operativo 40 que incluye procedimientos para manejar diversos servicios básicos del sistema y para realizar tareas dependientes del hardware;
- un módulo de comunicación opcional 42 que se usa para conectar el ordenador 10 a otros ordenadores a través de una o varias interfaces de comunicación 20 (cableadas o inalámbricas) y una o varias redes de comunicación 10 34, tales como Internet, otras redes de área amplia, redes de área local, redes del área metropolitana, y así sucesivamente;
- un módulo de interfaz de usuario opcional 42 que recibe comandos del usuario a través de los dispositivos de entrada 28, 72, etc. y genera objetos de interfaz de usuario en el dispositivo de visualización 26;
- 15 • un conjunto inicial de coordenadas tridimensionales $\{x_{1A_init}, \dots, x_{NA_init}, x_{1B_init}, \dots, x_{MB_init}, \dots, x_{PC_init}, \dots\}$ 46 para el polímero en estudio que comprende una pluralidad de átomos, en el que cada respectivo x_{iA} en $\{x_{1A_init}, \dots, x_{NA_init}, x_{1B_init}, \dots, x_{MB_init}, \dots, x_{PC_init}, \dots\}$ es una coordenada tridimensional para un átomo en un primer dominio en una pluralidad de dominios, cada respectivo x_{iB} en $\{x_{1A_init}, \dots, x_{NA_init}, x_{1B_init}, \dots, x_{MB_init}, \dots, x_{PC_init}, \dots\}$ es una coordenada tridimensional para un átomo en un segundo dominio en la pluralidad de dominios, cada respectivo x_{iC} en $\{x_{1A_init}, \dots, x_{NA_init}, x_{1B_init}, \dots, x_{MB_init}, \dots, x_{PC_init}, \dots\}$ es una coordenada tridimensional para un átomo en una primera bisagra del polímero, en el que el polímero está caracterizado por una capacidad para que el primer y el segundo dominio pivoten entre sí alrededor de la primera bisagra;
- 20 • un módulo de generación de polímero 50 que comprende instrucciones para alterar las coordenadas tridimensionales del polímero al pivotar el primer dominio con respecto al segundo dominio en torno a la primera bisagra, obteniendo así un conjunto alterado de coordenadas tridimensionales $\{x_{1A_init}, \dots, x_{NA_init}, x_{1B_init}, \dots, x_{MB_init}, \dots, x_{PC_init}, \dots\}$ para el polímero;
- 25 • un módulo de puntuación 52 para puntuar el conjunto alterado de coordenadas frente al criterio de Metropolis;
- una pluralidad de coordenadas tridimensionales alteradas 56 para el sistema molecular, donde comúnmente cada estructura alterada 56 tiene los mismos átomos que el sistema molecular en estudio, pero tiene coordenadas estructurales diferentes; y
- 30 • un módulo de refinamiento opcional 58 para refinar una o varias de las estructuras de polímero 56 alteradas usando dinámica molecular.

En algunas realizaciones, el polímero comprende entre 2 y 5.000 residuos, entre 20 y 50.000 residuos, más de 30 residuos, más de 50 residuos, o más de 100 residuos. En algunas realizaciones, un residuo en el polímero comprende 35 dos o más átomos, tres o más átomos, cuatro o más átomos, cinco o más átomos, seis o más átomos, siete o más átomos, ocho o más átomos, nueve o más átomos o diez o más átomos. En algunas realizaciones, el polímero 44 tiene un peso molecular de 100 dalton o más, 200 dalton o más, 300 dalton o más, 500 dalton o más, 1000 dalton o más, 5000 dalton o más, 10.000 dalton o más, 50.000 dalton o más o 100.000 dalton o más.

40 Un polímero, tales como aquellos que pueden ser estudiados usando los sistemas y métodos divulgados, es un gran sistema molecular compuesto de unidades estructurales que se repiten. Estas unidades estructurales que se repiten se denominan partículas o residuos de manera intercambiable en el presente documento. En algunas realizaciones, cada partícula p_i en el conjunto de $\{p_1, \dots, p_k\}$ partículas representa un único residuo diferente en el polímero nativo. Para ilustrar, considérese el caso donde el nativo comprende 100 residuos. En esta instancia, el conjunto de $\{p_1, \dots, p_k\}$ comprende 100 partículas, con cada partícula en $\{p_1, \dots, p_k\}$ representando una diferente de las 100 partículas.

En algunas realizaciones, el polímero que se evalúa usando los sistemas y métodos descritos es una proteína natural. En algunas realizaciones, el polímero es una proteína sintética.

50 En algunas realizaciones, el polímero es un heteropolímero (copolímero). Un copolímero es un polímero derivado de dos (o más) especies monoméricas, a diferencia de un homopolímero donde solo se usa un monómero. La copolimerización se refiere a los métodos usados para sintetizar químicamente un copolímero. Los ejemplos de copolímeros incluyen, pero no se limitan a, plástico ABS, SBR, caucho nitrilo, estireno-acrilonitrilo, estireno-isopreno-estireno (SIS) y etilvinilacetato. Dado que un copolímero consiste en al menos dos tipos de unidades constituyentes 55 (también unidades estructurales, o partículas), los copolímeros se pueden clasificar según la forma en que estas unidades se disponen a lo largo de la cadena. Estos incluyen copolímeros alternos con unidades regulares alternas A y B. Ver, por ejemplo, Jenkins, 1996, "Glossary of Basic Terms in Polymer Science" Pure Appl Chem. 68 (12): 2287-2311, que se incorpora en el presente documento por referencia en su totalidad. Los ejemplos adicionales de copolímeros son copolímeros periódicos con unidades A y B dispuestas en una secuencia de repetición (por ejemplo, (A-B-A-B-B-A-A-A-B-B-B)_n). Ejemplos adicionales de copolímeros son copolímeros estadísticos en los que la secuencia de residuos de monómeros en el copolímero sigue una regla estadística. Si la probabilidad de encontrar un residuo de tipo monómero dado en un punto particular de la cadena es igual a la fracción molar de ese residuo monomérico en la cadena, entonces el polímero puede ser referido como un copolímero verdaderamente aleatorio. Ver, por ejemplo, Painter, 1997, Fundamentals of Polymer Science, CRC Press, 1997, pág. 14, que se incorpora en el 60 presente documento por referencia en su totalidad. Otros ejemplos más de copolímeros que pueden evaluarse usando los sistemas y métodos divulgados son copolímeros de bloques que comprenden dos o más subunidades

homopolímeros enlazadas mediante enlaces covalentes. La unión de las subunidades de homopolímero puede requerir una subunidad intermedia que no se repite, conocida como un bloque de unión. Los copolímeros de bloque con dos o tres bloques distintos se llaman copolímeros dibloque y copolímeros tribloque, respectivamente.

En algunas realizaciones, el polímero es, de hecho, una pluralidad de polímeros, donde los polímeros respectivos en la pluralidad de polímeros no tienen todo el peso molecular. En tales realizaciones, los polímeros en la pluralidad de polímeros caen en un intervalo de peso con una distribución correspondiente de longitudes de cadena. En algunas realizaciones, el polímero es un sistema molecular de polímero ramificado que comprende una cadena principal con una o varias cadenas o ramas laterales sustituyentes. Los tipos de polímeros ramificados incluyen, pero no se limitan a, polímeros de estrella, polímeros de estrella, polímeros de cepillo, polímeros dendronizados, escaleras, y los dendrímeros. Ver, por ejemplo, Rubinstein y col., 2003, *Polymer physics*, Oxford; Nueva York: Oxford University Press, pág. 6, que se incorpora en el presente documento por referencia en su totalidad.

En algunas realizaciones, el polímero es un polipéptido. Como se usa en el presente documento, el término "polipéptido" significa dos o más aminoácidos o residuos enlazados por un enlace peptídico. Los términos "polipéptido" y "proteína" se usan indistintamente en el presente documento e incluyen oligopéptidos y péptidos. Un "aminoácido", "residuo" o "péptido" se refiere a cualquiera de las veinte unidades estructurales estándar de proteínas conocidas en la técnica, que incluyen iminoácidos, tales como prolina e hidroxiprolina. La designación de un isómero de aminoácido puede incluir D, L, R y S. La definición de aminoácido incluye aminoácidos no naturales. Por tanto, selenocisteína, pirrolisina, lantionina, ácido 2-aminoisobutírico, ácido gamma-aminobutírico, dehidroalanina, ornitina, citrulina y homocisteína se consideran aminoácidos. Otras variantes o análogos de los aminoácidos son conocidos en la técnica. Por tanto, un polipéptido puede incluir estructuras peptidomiméticas sintéticas tales como peptoides. Ver Simon y col., 1992, *Proceedings of the National Academy of Sciences USA*, 89, 9367, que se incorpora en el presente documento por referencia en su totalidad. Ver también Chin y col., 2003, *Science* 301, 964; y Chin y col., 2003, *Chemistry & Biology* 10, 511, cada uno de los cuales se incorpora por referencia en el presente documento en su totalidad.

Los polipéptidos evaluados de acuerdo con algunas realizaciones de los sistemas y métodos descritos también pueden tener cualquier número de modificaciones postraslacionales. Por tanto, un polipéptido incluye aquellos que son modificados por acilación, alquilación, amidación, biotilación, formilación, γ -carboxilación, glutamilación, glicosilación, glicilación, hidroxilación, yodación, isoprenilación, lipoilación, adición de cofactor (por ejemplo, de un hemo, flavina, metal, etc.), adición de nucleósidos y sus derivados, oxidación, reducción, pegilación, adición del fosfatidilinositol, fosfopantoetilación, fosforilación, formación de piroglutamato, racemización, adición de aminoácidos por ARNt (por ejemplo, arginilación), sulfatación, selenoilación, ISGylation, SUMOylation, ubiquitinación, modificaciones químicas (por ejemplo, citrulinación y deamidación), y tratamiento con otras enzimas (por ejemplo, proteasas, fosfatasas y quinasas). Otros tipos de modificaciones postraslacionales se conocen en la técnica y también se incluyen.

Se divulga que el polímero es un complejo organometálico. Un complejo organometálico es un compuesto químico que contiene enlaces entre el carbono y el metal. En ciertas instancias, los compuestos organometálicos se distinguen por el prefijo "organo-", por ejemplo, compuestos de organopalladio. Los ejemplos de tales compuestos organometálicos incluyen todos los reactivos de Gilman, que contienen litio y cobre. Tetracarbonilo de níquel, y ferroceno son ejemplos de compuestos organometálicos que contienen metales de transición. Otros ejemplos incluyen compuestos de organomagnesio como yodo(metil)magnesio MeMgI, dietilmagnesio (Et₂Mg), y todos los reactivos de Grignard; compuestos de organolitio tales como n-butilitio (n-BuLi), compuestos de organozinc tales como dietilzinc (Et₂Zn) y cloro (etoxicarbonilmetilo)zinc (ClZnCH₂C(=O)OEt); y compuestos organocúpricos tales como dimetilcuprato de litio (Li⁺[CuMe₂]⁻). Además de los metales tradicionales, lantánidos, actínidos, y semimetales, elementos como el boro, silicio, arsénico, y selenio se consideran compuestos organometálicos en forma, por ejemplo, compuestos de organoborano tales como trietilborano (Et₃B).

En algunas realizaciones, el conjunto inicial de coordenadas tridimensionales 46 para el polímero se obtiene por cristalografía de rayos X, técnicas espectroscópicas de resonancia magnética nuclear, o microscopía electrónica. En algunas realizaciones, el conjunto de M coordenadas tridimensionales {x₁, ..., x_M} se obtiene por modelado (por ejemplo, simulaciones de dinámica molecular).

En algunas realizaciones, el polímero incluye dos tipos diferentes de polímeros. En algunas realizaciones, el polímero incluye dos polipéptidos enlazados entre sí. En algunas realizaciones, el polímero en estudio incluye uno o varios iones metálicos (por ejemplo, una metaloproteína con uno o más átomos de zinc) y/o está enlazado a una o varias moléculas orgánicas pequeñas (por ejemplo, un inhibidor). En tales instancias, los iones metálicos y/o las moléculas pequeñas orgánicas pueden representarse como una o varias partículas adicionales p_i en el conjunto de {p₁, ..., p_k} partículas que representan el polímero nativo.

El polímero en estudio es una proteína. Los elementos estructurales básicos de las proteínas son bien conocidos en la técnica. Los aminoácidos no terminales comúnmente tienen la estructura -NH-C^αHR-CO-, donde R representa una cadena lateral de aminoácido como se conoce en la técnica. Los átomos tales como N, C^α, C^o y O que no están en la cadena lateral representan átomos de la red troncal. Los átomos de la cadena lateral, especialmente los heteroátomos de la cadena lateral, son denominados como átomos "terminales". Por tanto, los átomos terminales incluyen C^β en alanina, S^γ en cisteína, y N^{ε1} y C^{η1} en triptófano, por ejemplo. Tales átomos terminales pueden ser únicos. C-alfa o C^α

es el átomo de carbono en el centro de cada aminoácido. La red troncal de la proteína incluye N, C-alfa, C y O átomos. Los ángulos diedros de la red troncal de las proteínas se llaman ϕ (phi, involucrando los átomos de la red troncal C'-N-C α -C'), ψ (psi, involucrando los átomos de la red troncal N-C α -C'-N) y ω (omega, involucrando los átomos de la red troncal C α -C'-N-C α). Por tanto, ϕ controla la distancia C'-C', ψ controla la distancia N-N y ω controla la distancia C α -C α .

La planaridad del enlace peptídico generalmente restringe que ω sea 180° (el caso trans típico) o 0° (el caso cis raro). Los ángulos diedros de cadena lateral tienden a agruparse cerca de 180°, 60°, y -60°, que se llaman los trans, gauche⁺, y las conformaciones gauches. La elección de los ángulos diedros de cadena lateral se ve afectada por la red troncal vecina y los diedros de cadena lateral. Un mapa de Ramachandran (Ramachandran, Ramakrishnan, y Sasisekharan 1963) es una representación de las geometrías de la red troncal de la proteína estereoquímicamente permitidas en función de sus ángulos de torsión variables. El método permite la clasificación de las conformaciones de la red troncal de la proteína como permitidas, parcialmente permitidas o no permitidas basándose en estos parámetros geométricos.

Hay diferentes niveles de descripción de la estructura de una proteína. La estructura primaria se refiere a la secuencia lineal de aminoácidos que forman la cadena polipeptídica. El enlace entre dos aminoácidos es un enlace peptídico. La secuencia de aminoácidos determina el posicionamiento de los diferentes grupos R entre sí. Este posicionamiento determina la forma en que la proteína se pliega y la estructura final de la molécula. La estructura secundaria de las moléculas de proteína se refiere a la formación de un patrón regular de giros o torceduras de la cadena polipeptídica. La regularidad se debe a la formación de enlaces de hidrógeno entre los átomos de la red troncal de aminoácidos de la cadena polipeptídica. Los dos tipos más comunes de estructura secundaria se denominan "hélice α " y "hoja plegada β ". La estructura terciaria se refiere a la estructura globular tridimensional formada por la flexión y retorcimiento de la cadena polipeptídica. Este proceso a menudo significa que la secuencia lineal de aminoácidos se pliega en una estructura globular compacta. El plegado de la cadena polipeptídica se estabiliza por múltiples interacciones débiles no covalentes. Estas interacciones incluyen enlaces de hidrógeno, interacciones electrostáticas, interacciones hidrofóbicas, y en ocasiones, enlaces covalentes. La estructura cuaternaria se refiere al hecho de que algunas proteínas contienen más de una cadena polipeptídica, añadiendo un nivel adicional de organización estructural: la asociación de las cadenas polipeptídicas. Cada cadena polipeptídica en la proteína se llama subunidad. Las subunidades pueden ser la misma cadena polipeptídica o diferentes. Por ejemplo, la enzima β -galactosidasa es un tetrámero, lo que significa que está compuesta por cuatro subunidades, y, en este caso, las subunidades son idénticas: cada cadena polipeptídica tiene la misma secuencia de aminoácidos. La hemoglobina, la proteína que lleva el oxígeno en la sangre, también es un tetrámero, pero está compuesto por dos cadenas polipeptídicas de un tipo (141 aminoácidos) y dos de un tipo diferente (146 aminoácidos).

Generalmente se entiende que un polímero exhibe cierto grado de movilidad. Como resultado, en cualquier punto dado, un polímero puede existir en un número de "conformaciones", "estados conformacionales" o "conformadores", es decir, en una de un número de configuraciones geométricas diferentes que tienen la misma fórmula o conectividad estructural. El término "conformación alterada", según se usa en el presente documento, se refiere a una conformación que resulta de realizar los métodos descritos aquí en una conformación de inicio o entrada. El polímero es una proteína y una conformación comprende un rotámero de un aminoácido, mientras que otra conformación comprende un rotámero diferente del mismo aminoácido. El término "rotámero" se refiere a una combinación de ángulos de torsión que describen la geometría de una cadena lateral de aminoácidos.

Los métodos divulgados también pueden usarse para modelar un "sistema", que comprende un polímero y puede comprender, además, otras entidades moleculares tales como disolventes, iones y similares. El modelado del propio polímero podría ser, sin limitación, atómica o de grano grueso.

En algunas realizaciones, los programas o módulos identificados en la figura 1 corresponden a conjuntos de instrucciones para llevar a cabo una función descrita en este el presente documento. Los conjuntos de instrucciones pueden ejecutarse por uno o varios procesadores (por ejemplo, las CPU 22). Los módulos o programas identificados anteriormente (por ejemplo, conjuntos de instrucciones) no necesitan implementarse como programas de software, procedimientos o módulos separados, y, por lo tanto, varios subconjuntos de estos programas o módulos pueden combinarse o reorganizarse de otra manera en diversas realizaciones. En algunas realizaciones, la memoria 36 almacena un subconjunto de los módulos y estructuras de datos identificados anteriormente. Asimismo, la memoria 36 puede almacenar módulos adicionales y estructuras de datos no descritas anteriormente.

Ahora que se ha descrito un sistema de acuerdo con los sistemas y métodos de la presente divulgación, la atención se dirige a la figura 2, que ilustra un método a modo de ejemplo de acuerdo con la presente divulgación.

Etapa 402. Los residuos de un polímero pueden estar caracterizados de acuerdo con su grado de movilidad. Por ejemplo, ciertos aminoácidos, incluyendo los que se encuentran en motivos estructurales secundarios, como las hélices α y las hojas β , pueden estar caracterizados por ser rígidos, mientras que otros aminoácidos, tales como los de los bucles que conectan estos motivos, puede estar caracterizados como siendo móviles.

Los movimientos a gran escala en un polímero a veces se pueden clasificar como un movimiento de bisagra. Un "movimiento de bisagra" implica la rotación de partes de proteína (por ejemplo, dominios) sobre una "bisagra" o una "región de bisagra", que se refieren a uno o varios residuos móviles. Los residuos móviles tienden a exhibir una distribución más amplia de conformaciones en un conjunto de conformaciones de polímeros, mientras que los residuos

rígidos tienden a exhibir una distribución más estrecha.

Por consiguiente, en la etapa 402, un conjunto inicial de coordenadas tridimensionales $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$ para un polímero se obtiene, donde el polímero comprende una pluralidad de dominios, cada respectivo x_{iA} en $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$ es una coordenada tridimensional para un átomo en un primer dominio en la pluralidad de dominios, cada respectivo x_{iB} en $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$ es una coordenada tridimensional para un átomo en un segundo dominio en la pluralidad de dominios, y cada respectivo x_{iC} en $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$ es una coordenada tridimensional para un átomo en una primera bisagra del polímero, en donde el polímero está caracterizado por una capacidad para que el primer y el segundo dominio pivoten entre sí alrededor de la primera bisagra. Esto representa las características mínimas del polímero. El polímero puede comprender cualquier número de dominios y cualquier número de bisagras.

Una bisagra generalmente involucra varios residuos que experimentan cambios conformacionales simultáneos significativos, mientras que los residuos de las porciones giratorias del polímero que no están en la bisagra permanecen sin cambios con respecto a los otros aminoácidos en sus respectivos dominios. En un polímero, puede haber puntos de cambio direccional a lo largo de la secuencia primaria que resultan en la geometría interdominio necesaria para la estructura o actividad enzimática. Forman regiones de restricción relativa en términos de movimiento molecular. En muchas instancias, un bucle, un enlazador o una porción del mismo puede estar caracterizado como una "bisagra" o una "región de bisagra".

En algunas realizaciones, una bisagra comprende residuos que son relativamente distantes en secuencia, pero cercanos en espacio. Por ejemplo, la estructura del anticuerpo IgG (figura 4) está compuesta por una estructura homodimérica, estando cada una de las unidades monoméricas compuesta por una cadena pesada con cuatro dominios y una cadena ligera con dos dominios. La estructura de la IgG dimérica se clasifica ampliamente en una sección Fe (fragmento cristizable) y dos secciones Fab (unión de antígeno de fragmento). Cada una de las secciones Fab está conectada a la Fe por un enlazador de quince aminoácidos y, este enlazador se conoce tradicionalmente como la bisagra en la estructura Ab. Se cree que la flexibilidad lograda en la estructura de IgG debido a la bisagra es importante para la eficacia funcional de la molécula Ab. Como otro ejemplo, se puede considerar que la proteína de unión a glutamina (ID PDB 1 ggg y 1 wdn) tiene dos partes rígidas que están conectadas por dos residuos de bisagra 87 y 183. Mientras se separa secuencialmente por 95 residuos, los átomos C_{α} de los residuos 87 y 183 están separados aproximadamente 6 Å. Por tanto, en algunas realizaciones, una región de bisagra comprende un conjunto de puntos finales agrupados espacialmente de una porción rígida de un polímero. Las encuestas de estructuras resueltas cristalográficamente que evidencian movimientos de dominio en proteínas han revelado que los movimientos de bisagra y corte son elementos básicos que producen los movimientos. Ver gerstein, Lesk, y Chothia, 1994, "Structural mechanisms for domain movements in proteins", Biochemistry 33 (22):6739-49, que se incorpora aquí por referencia.

Las bisagras de un polímero también pueden determinarse mediante un análisis seleccionado del grupo que consiste en análisis de modo normal, análisis de componentes principales, análisis de covarianza, análisis de dominio de proteína, análisis de rigidez y análisis de característica de secuencia. En algunas realizaciones, el análisis de modo normal de un modelo de grano grueso se emplea en la predicción de las bisagras. En alguna realización, la bisagra se puede definir basándose en la accesibilidad del solvente y los contactos realizados por los residuos. Los residuos de las bisagras suelen ser accesibles a los solventes y hacen un contacto mínimo con el resto de los dominios de proteínas.

El análisis de modo normal (NMA) examina las propiedades vibracionales y térmicas de las estructuras moleculares a nivel atómico. Ver Bahar y Rader, 2005, "Coarse-grained normal mode analysis in structural biology", Curr Opin Struct Biol 15(5): 586-592, que se incorpora aquí por referencia. Este método proporciona soluciones analíticas para describir los principales modos de movimiento, también conocidos como "modos normales", para una estructura alrededor de su estado de equilibrio, caracterizados por las diferencias de frecuencia. Dado que el método describe el movimiento de la proteína alrededor de una posición de equilibrio, no puede proporcionar una visión directa de los diversos subestados conformacionales de la proteína, o las transiciones entre ellos, que la dinámica molecular o las simulaciones de Monte Carlo pueden. Sin embargo, el método proporciona información sobre las características de movimiento de baja frecuencia del sistema, información que se puede usar para definir dominios de polímeros y las bisagras que conectan los dominios en la estructura del polímero. Realizar NMA en moléculas más grandes es un desafío computacional porque el movimiento de equilibrio de la molécula debe describirse mediante una matriz constante de fuerza derivada de la segunda derivada del potencial con respecto a las coordenadas cartesianas.

Los modelos de Red Elástica tratan el polímero como en NMA, pero describen el polímero como un conjunto de partículas que están interconectadas por una red de resortes elásticos. Ver Bahar, Atilgan, y Erman, 1997, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential", Fold Des 2 (3):173-81, que se incorpora aquí por referencia. Los modelos empleados son más simples que los modelos empleados en NMA. Dependiendo de la granularidad del modelo, las partículas pueden corresponder a átomos en la proteína o puntos representativos de grano grueso, como las posiciones del átomo C-alfa en la cadena de proteína, posiciones de cadenas laterales o definiciones de grupos de átomos alternativos. Los modelos pueden emplearse para describir las fluctuaciones estructurales en torno a una conformación estable de una proteína. Estas fluctuaciones son isotrópicas

en un Modelo de Red Gaussiana (ver Haliloglu y col., 1997, "Gaussian Dynamics of Folded Proteins", Physical Review Letters 79 (16):3090, que se incorpora aquí por referencia en el presente documento en su totalidad) y anisotrópicas en el caso del Modelo de Red Anisotrópica (ver Atilgan y col., 2001, "Anisotropy of fluctuation dynamics of proteins with an elastic network model", Biophys J 80 (1):505-15, que se incorpora en el presente documento por referencia en su totalidad).

El análisis de componentes principales (PCA) es un procedimiento matemático que reduce un número de variables correlacionadas a menos variables no correlacionadas llamadas "componentes principales". El primer componente principal se selecciona de manera que tenga en cuenta la mayor variabilidad posible en los datos, y cada componente subsiguiente da cuenta de la mayor variabilidad posible. El propósito de PCA es descubrir o reducir la dimensionalidad del conjunto de datos, e identificar nuevas variables subyacentes significativas.

La PCA se logra mediante el establecimiento de datos reales en una matriz de covarianza o una matriz de correlación. La técnica matemática usada en PCA se llama análisis Eigen: uno resuelve los valores propios y los vectores propios de una matriz simétrica cuadrada con sumas de cuadrados y productos cruzados. El vector propio asociado al valor propio más grande tiene la misma dirección que el primer componente principal. El vector propio asociado al segundo valor propio más grande determina la dirección del segundo componente principal. La suma de los valores propios es igual a la traza de la matriz cuadrada y el número máximo de vectores propios es igual al número de filas (o columnas) de esta matriz.

La información de covarianza en algunas realizaciones se puede derivar del cálculo de un mapa dinámico de correlación cruzada (DCCM). Un DCCM es un mapa de coeficientes de correlación cruzada para el desplazamiento de cualquiera de los dos átomos i y j , y está dado por:

$$C_{ij} = \langle \Delta r_i \bullet \Delta r_j \rangle / \sqrt{\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle}$$

donde Δr_i es el desplazamiento del átomo i de su posición media. El mapa y los coeficientes de correlación correspondientes pueden resaltar regiones de proteína que se mueven en concierto de forma correlacionada con respecto a regiones estructurales que están correlacionadas en su dinámica. Por tanto, en algunas realizaciones, la covarianza de los desplazamientos de residuos en la estructura de una proteína se usa para identificar las regiones de bisagra en la proteína.

En algunas realizaciones, la conformación alterada está desviada hacia un modo de movimiento de baja frecuencia principal determinado por un análisis seleccionado del grupo que consiste en el análisis de componentes principales de la trayectoria de dinámica molecular y el análisis de modelos de redes elásticas anisotrópicas.

Etapa 404. En la etapa 404, el conjunto de coordenadas tridimensionales del polímero se altera pivotando el primer dominio con respecto al segundo dominio sobre la primera bisagra, obteniendo así un conjunto alterado de coordenadas tridimensionales $\{X_{1A_alt}, \dots, X_{NA_alt}, X_{1B_alt}, \dots, X_{MB_alt}, \dots, X_{PC_alt}, \dots\}$ para el polímero, donde los átomos dentro del primer dominio se mantienen fijos entre sí y los átomos dentro del segundo dominio se mantienen fijos entre sí.

Para apreciar la etapa 404, es instructivo señalar que uno o varios parámetros de bisagra están asociados a cualquier bisagra particular. Un parámetro de bisagra se refiere a cualquier cantidad que se puede usar para describir o definir un aspecto físico de una bisagra. Por ejemplo, los parámetros de bisagra útiles incluyen, sin limitación, ángulos de torsión, ángulos de enlace y distancias interatómicas. En algunas realizaciones, el parámetro bisagra es un ángulo de torsión. En realizaciones a modo de ejemplo, el parámetro de bisagra es un ángulo de torsión seleccionado del grupo que consiste en ángulos diedros ϕ y ψ .

En algunas realizaciones, el polímero comprende una pluralidad de bisagras caracterizadas por una pluralidad de parámetros de bisagra. La pluralidad de parámetros de bisagra puede comprender solo un tipo de parámetro de bisagra o cualquier combinación de diferentes parámetros de bisagra.

Algunas realizaciones de la etapa 404 comprenden aplicar una alteración a uno o varios parámetros de bisagra de la primera bisagra para pivotar uno de los dos dominios conectados a la primera articulación con respecto al otro de los dos dominios. En algunas realizaciones, la alteración se aplica a un subconjunto de los parámetros de bisagra de la primera bisagra. En algunas realizaciones, la alteración se aplica a un subconjunto de una pluralidad de parámetros de bisagra, donde el subconjunto es elegido al aleatoriamente. En algunas realizaciones, la alteración es aleatoria.

En algunas realizaciones, una alteración de un parámetro de bisagra que se aplica en la etapa 404 para lograr el pivote comprende una rotación en torno a de un solo ángulo diedro o rotación concertada en torno a dos o más ángulos diedros en posiciones distantes a lo largo de la secuencia de polímero. En estas realizaciones, se puede capturar un movimiento de interdominio abisagrado o movimiento de bucle. En algunas realizaciones, movimientos de espalda (Davis y col. 2006) se emplean para introducir la alteración de la bisagra. Ver Davis, 2006, "The backrub motion: How protein backbone shrugs when a sidechain dances", Structure 14:265-274, que se incorpora en el presente documento

por referencia en su totalidad, para una descripción de esta forma de movimiento.

En algunas realizaciones en las que el polímero es una proteína, las geometrías de la cadena lateral se empaquetan en conformaciones alternativas una vez que la bisagra ha pivotado mediante la aplicación de un algoritmo de empaquetamiento de proteínas. Se ha desarrollado un número de algoritmos de empaquetamiento de proteínas para buscar y optimizar las geometrías de las cadenas laterales de aminoácidos a partir de una geometría de red troncal de la proteína. Ver, por ejemplo, Jones, 1994, "De novo protein design using pairwise potentials and a genetic algorithm" *Protein Sci* 3 (4):567-74; Desmet y col, 1992, "The dead-end elimination theorem and its use in protein side-chain positioning" *Nature* 356 (6369):539-542; Holm y Sander, 1992, "Fast and simple Monte Carlo algorithm for side chain optimization in proteins: application to model building by homology" *Proteins* 14 (2):213-23; Lee y Subbiah, 1991, "Predicción de la conformación de la cadena lateral de la proteína mediante la optimización del empaque", *J Mol Biol* 217 (2):373-88; y Kono y Doi, 1994, "Energy minimization method using automata network for sequence and side-chain conformation prediction from given backbone geometry", *Proteins* 19 (3):244-55, cada uno de los cuales se incorpora en el presente documento en su totalidad. Dada una serie de geometrías de cadenas laterales de aminoácidos conocidas como rotámeros, estos algoritmos resuelven la combinación de la geometría óptima del rotámero para todos los aminoácidos en la secuencia de proteína. En algunas realizaciones, la etapa de reempaquetado comprende la optimización de la geometría del rotámero de cadena lateral para los aminoácidos que se ven afectados por la alteración, según se describe en el presente documento.

En algunas realizaciones de la etapa 404, el primer dominio pivota con respecto al segundo dominio en cierta cantidad sin la aplicación de un cambio a un parámetro de bisagra. De hecho, en tales realizaciones, los residuos de la primera bisagra se descartan por completo. Una vez realizado el pivote, la primera bisagra se reconstruye de nuevo basándose en las nuevas ubicaciones del primer y del segundo dominio. Se revela que dicha reconstrucción se puede lograr mediante la implementación del "problema de cierre de bucle". El problema de cierre de bucle ha sido aplicado, por ejemplo, en la predicción de geometría de bucle en modelos de homología de proteínas o mediante métodos conocidos de predicción de la estructura de secuencias de bucle faltantes en estructuras de proteínas. El problema de cierre de bucle en el contexto de estructuras de proteínas también se conoce como cinemática inversa en el campo matemático. Dada una cadena de polímero tal que los puntos finales de la cadena se limitan a ubicaciones específicas, el problema se relaciona con encontrar los valores correctos para los grados de libertad subyacentes, por ejemplo, la red troncal de los ángulos diedros en la cadena del polímero. Existen varias soluciones analíticas para este problema. Ver, por ejemplo, Wedemeyer y Scheraga, 1999, "Exact analytical loop closure in proteins using polynomial equations", *Journal of Computational Chemistry* 20 (8):819-844; Coutsiyas y col., 2004, "A kinematic view of loop closure", *J Comput Chem* 25 (4):510-28, cada uno de los cuales se incorpora aquí por referencia. Se revela que los métodos comprenden resolver el problema de cierre de bucle asociado a múltiples cambios de ángulo de torsión distantes realizados individualmente.

Etapla 406. En la etapa 406 se calcula un valor de energía para el conjunto alterado de coordenadas producido por la última instancia de la etapa 404.

El término "subestado estructural" describe un estado energético particular. Un polímero plegado con una estructura particular ocupa un mínimo de energía potencial. Sin embargo, a menudo hay más de un mínimo "local" de energía similarmente baja. Cuando un polímero ocupa uno cualquiera de estos mínimos, tiene una estructura distinta, o "subestado". Además de la interconversión entre subestados, los polímeros experimentan fluctuaciones estructurales continuas dentro de un subestado particular mínimo. La puntuación de la etapa 406 ayuda a identificar tales subestados.

En algunas realizaciones, puntuar la conformación alterada comprende puntuar una métrica de rendimiento con un criterio de Metropolis. Se divulga que en la técnica se conocen varias métricas de rendimiento. En algunas realizaciones, la métrica de rendimiento es una función de energía potencial. Una función de energía potencial es una función matemática que comprende los términos que representan los enlaces, ángulos, diedros impropios, torsión e interacciones no enlazadas en un polímero. Una función de energía potencial comprende cualquier combinación o subconjunto de estos términos, y también puede incluir términos adicionales. Tal función de energía puede denominarse "función de energía basada en la física". Se divulga que la métrica de rendimiento es una función de energía basada en el conocimiento, que es una forma alternativa de función de energía potencial en la que los parámetros se derivan de un análisis estadístico de frecuencias de par atómico o de residuos observados en estructuras tridimensionales de polímeros solucionadas experimentalmente.

Las interacciones no enlazadas se refieren a secciones de la función de energía potencial que se tratan a través de las interacciones espaciales entre los átomos en el sistema que no interactúan a través de enlaces, ángulos y términos diedros. Estos incluyen interacciones con solventes y otras moléculas ambientales. Como una aproximación, en función de la separación entre los átomos, pueden ser factorizados o ignorados en un modelo dado. Se divulga que la métrica de rendimiento es una función de energía potencial que comprende un término de valencia, un término de interacción no enlazado, un término de efecto de solvatación o cualquier combinación de los mismos.

Etapla 408. En la etapa 408, se determina si el criterio de Metrópolis se considera satisfecho por la métrica de rendimiento. Si no es así (408-No), el control del proceso regresa a la etapa 404 donde las coordenadas del polímero

se alteran nuevamente al pivotar. Si es así (408-Si), el control del proceso pasa a la etapa opcional 410 o el proceso finaliza.

En realizaciones a modo de ejemplo, el bucle 404-408 constituye una búsqueda de Monte Carlo en la conformación de polímeros. Los métodos de Monte Carlo son una clase de algoritmos computacionales que se basan en muestreos aleatorios repetidos para calcular los resultados. Tales algoritmos son útiles en situaciones en las que no es práctico calcular un resultado exacto con un algoritmo determinista. La simulación de Monte Carlo usa un criterio de Metrópolis. El criterio de Metropolis en una simulación de Monte Carlo introduce una función de energía dependiente de la temperatura que sigue un balance detallado para lograr un muestreo de estados en equilibrio. Ver Metropolis y col., 1953, "Equation of State Calculations by Fast Computing Machines", The Journal of Chemical Physics 21 (6):1087-1092, que se incorpora aquí por referencia.

En los enfoques tradicionales de Monte Carlo, las perturbaciones se introducen en cualquier lugar a lo largo de la secuencia primaria del polímero y se limitan a magnitudes muy pequeñas. En estos métodos tradicionales, hacer grandes movimientos lleva a, en la mayoría de los casos, una perturbación de las estructuras de dominio interno y estados energéticos desfavorables. Asimismo, en el caso de los polímeros que son proteínas, estas perturbaciones se introducen comúnmente en los $2(n-2)$ ángulos diedros de la red troncal, donde n es el número de la pluralidad de aminoácidos de la proteína. La presente divulgación proporciona, en algunas realizaciones, métodos que utilizan perturbaciones de parámetros de bisagra que tienen más probabilidades de conducir a estados energéticamente factibles, y que serán aceptados por el algoritmo. En algunas realizaciones, las perturbaciones están restringidas a un subconjunto de ángulos diedros en la bisagra. En realizaciones a modo de ejemplo, se modifican menos de $2(n-2)$ parámetros de bisagra, donde n es el número de residuos en el polímero. En estas realizaciones, el número reducido de parámetros de bisagra perturbados da como resultado ventajosamente un aumento de la velocidad y la eficacia en la ejecución del método. Por tanto, los métodos divulgados superan un obstáculo importante en el uso de los enfoques tradicionales de Monte Carlo para el muestreo de conformaciones de polímeros.

En algunas realizaciones, los métodos comprenden aceptar o rechazar una conformación alterada en función de un criterio de Metropolis, como se mencionó anteriormente. En otras palabras, las instancias respectivas de las etapas 404 y 406 se realizan secuencialmente en el tiempo y una enésima instancia de la etapa 404 altera el conjunto alterado de coordenadas del polímero producido por la instancia enésima-1 o enésima-2 de la etapa de modificación 404, donde esta selección del conjunto alterado de coordenadas del polímero para la enésima instancia de la etapa de alteración 404 se basa en los criterios de Metropolis, y donde n es un número entero positivo de tres o mayor. Por tanto, si no se satisfacen los criterios de Metropolis, la enésima instancia de la etapa 404 de alteración utiliza las coordenadas para la enésima-2 instancia de la etapa 404 de modificación y si se satisface el criterio de Metropolis, la enésima instancia de la etapa de modificación 404 usa las coordenadas para la enésima-1 instancia de la etapa de alteración 404.

En algunas realizaciones, el parámetro de bisagra es un ángulo de torsión de red troncal y la conformación alterada está desviada hacia estados diedros estereoquímicamente aceptables. La técnica proporciona numerosas formas de determinar estados diedros estereoquímicamente aceptables, por ejemplo, a través de los mapas de Ramachandran (tratados anteriormente) y similares.

Etapas 410. Opcionalmente, el conjunto alterado de coordenadas sirve como punto de partida para una simulación atómica. Una simulación atómica se refiere a un protocolo de simulación basado en un modelo que describe un sistema molecular en términos de los átomos individuales presentes en el sistema. Los términos de valencia que describen la topología de enlace que conecta los átomos junto con las interacciones atómicas por pares no enlazados se emplean para describir la energía del sistema. Ver, Adcock McCammon, 2006, "Molecular dynamics: survey of methods for simulating the activity of proteins". Chem Rev 106 (5):1589-615. Los "modelos de grano grueso" simplifican o abstraen algunos de los aspectos del modelo molecular para obtener alguna ventaja, por ejemplo, para reducir la necesidad de recursos informáticos.

Relativo al modelado de grano grueso, el modelado de grano fino o atómico aumenta el nivel de detalle en el sistema molecular que se está modelando. En algunas realizaciones, un conjunto alterado de coordenadas tridimensionales del polímero se somete a grano fino, simulaciones de dinámica molecular de todos los átomos. De este modo, es más probable que una simulación de dinámica molecular conduzca al muestreo de subestados conformacionales únicos.

Por tanto, en realizaciones a modo de ejemplo, los métodos comprenden llevar a cabo una simulación de dinámica molecular de una conformación alterada. La dinámica molecular (MD) es una forma de simulación informática en la que los átomos y las moléculas se modelan en interacción a lo largo del tiempo utilizando funciones de energía potenciales (como se describió anteriormente) que se aproximan a las propiedades físicas conocidas. Los sistemas moleculares comprenden un gran número de partículas, haciendo imposible establecer la naturaleza de los sistemas analíticamente. La simulación de MD evita este problema mediante el uso de métodos numéricos y permite a los investigadores caracterizar las propiedades dinámicas del sistema.

En una descripción simplificada del algoritmo de simulación de dinámica molecular, la simulación se realiza de forma iterativa, calculando alternativamente las fuerzas y resolviendo las ecuaciones de movimiento basadas en las aceleraciones obtenidas de las nuevas fuerzas. De hecho, sin estar limitado por la teoría, la mayoría de las prácticas

MD usan versiones mucho más complicadas del algoritmo, incluyendo dos etapas para resolver las ecuaciones de movimiento y muchas etapas adicionales que explican el control de la temperatura y la presión, análisis y salida. Numerosos métodos de simulación de dinámica molecular son conocidos en la técnica y se pueden usar para realizar el bucle 404-408. Adcock y McCammon (2006) proporcionan una revisión general de los métodos de dinámica molecular útiles en la presente invención. Ver, Adcock y McCammon, 2006, "Molecular dynamics: survey of methods for simulating the activity of proteins". Chem Rev 106 (5):1589-615.

En algunas realizaciones, la simulación de dinámica molecular comprende el uso de un modelo de solvente implícito o explícito. En contraste con los modelos explícitos de solventes en los cuales el solvente es tratado a nivel atómico, los modelos de solventes implícitos representan el solvente como un continuo. Ver, por ejemplo, Lazaridis y Karplus, 1999, "Effective energy function for proteins in solution", Proteins: Structure, Function, and Genetics 35: 133-152, que se incorpora en el presente documento por referencia. En algunas realizaciones, los métodos comprenden calcular la energía de solvatación de un sistema mediante un método de solvente implícito.

En algunas realizaciones, la simulación de dinámica molecular comprende el cálculo de un panorama energético libre basado en la conformación alterada mediante el uso de cualquiera de un número de métodos adecuados. En realizaciones a modo de ejemplo, un panorama energético libre basado en la conformación alterada se calcula mediante un método seleccionado de Muestreo de Paraguas, Integración Termodinámica, Perturbación de Energía Libre, Fuerza de Polarización Adaptativa basada en el potencial de los cálculos de fuerza media y dinámica molecular dirigida. Muestreo de Paraguas, Integración Termodinámica, Perturbación de Energía Libre, o Fuerza de Polarización Adaptativa basada en el potencial de los cálculos de fuerza son métodos de simulación de energía libre que emplea mecanismos estadísticos basados en el enfoque para derivar diferencia de energía libre entre dos estados de un sistema.

En algunas realizaciones, los métodos comprenden llevar a cabo una simulación de dinámica molecular de Intercambio de Réplica en Serie (SRE). El intercambio de réplica se refiere a un algoritmo de simulación molecular en el que se lleva a cabo un número de simulaciones de dinámica molecular del sistema de interés a diferentes temperaturas de forma sincrónica. El método de Intercambio de Réplica en Serie es un método asíncrono para ejecutar simulaciones de un sistema a diferentes temperaturas. Ver, por ejemplo, Hagen y col., 2007, "Intercambio de Réplicas en Serie". J Phys Chem B 111 (6):1416-23, que se incorpora en el presente documento por referencia en su totalidad. En algunas realizaciones, la simulación de dinámica molecular SRE comprende templado de temperatura o templado de solvente.

En algunas realizaciones, los métodos comprenden generar una pluralidad de conformaciones alteradas. Es decir, se guardan algunos de los conjuntos de coordenadas tridimensionales alteradas del polímero de varias instancias de la etapa 404. En algunas realizaciones, se guarda cada conjunto de coordenadas tridimensionales alteradas que satisface alguna forma de función de evaluación predeterminada. En algunas realizaciones, esta función de evaluación es la misma que la métrica de rendimiento evaluada por el criterio de Metropolis solo con un criterio menos estricto para la satisfacción de la función de evaluación. En algunas realizaciones, el estado de protonación de una cadena lateral de aminoácido ionizable varía entre la pluralidad de conformaciones alteradas. En algunas realizaciones, los métodos comprenden asignar cada una de la pluralidad de conformaciones alteradas a uno de una pluralidad de clústeres conformacionales, donde la etapa de asignación comprende el uso de datos estructurales seleccionados del grupo que consiste en un parámetro geométrico interdominio, un patrón de contacto interatómico, un radio de rotación, la accesibilidad del solvente y la desviación cuadrática media de la pluralidad de conformaciones alteradas. El proceso de separar una pluralidad de conformaciones alteradas en diversos subconjuntos ("clústeres conformacionales") se conoce como "agrupamiento". El criterio para la agrupación se define de acuerdo con las necesidades del profesional. Por ejemplo, las conformaciones alteradas de un clúster conformacional pueden compartir valores de una propiedad seleccionada del grupo que consiste en la geometría interdominio, criterios de empaquetado, tales como orden de contacto, accesibilidad de los residuos de "punto caliente", y similares. La pluralidad de conformaciones alteradas de un clúster conformacional también puede agruparse de acuerdo con los datos estructurales seleccionados del grupo que consiste en un parámetro geométrico interdominio, un patrón de contacto interatómico, un radio de rotación, la accesibilidad del solvente y la desviación cuadrática media de la pluralidad de conformaciones alteradas. Las conformaciones agrupadas pueden analizarse adicionalmente mediante la aplicación de métodos tales como la dinámica molecular.

En algunas realizaciones, los métodos comprenden realizar una pluralidad de simulaciones de dinámica molecular independientes. En algunas realizaciones, la pluralidad de simulaciones de dinámica molecular independientes se realiza en paralelo (a) en nodos de un clúster de informático, (b) en un sistema informático distribuido o (c) mediante computación de propósito general en unidades de procesamiento de gráficos. En algunas realizaciones, si una similitud conformacional entre estructuras intermedias en cualquiera de los dos de la pluralidad de simulaciones de dinámica molecular independiente está dentro de un cierto umbral, una de las dos simulaciones de dinámica molecular se detiene. En algunas realizaciones, la pluralidad de simulaciones de dinámica molecular independientes se realiza en estructuras representativas de clústeres conformacionales descritos anteriormente.

En algunas realizaciones, la simulación de dinámica molecular comprende el cálculo adaptativo de una distribución de energía de una pluralidad de conformaciones.

Implementación en un sistema informático. Los métodos descritos pueden implementarse como programas informáticos que se ejecutan en ordenadores programables que comprenden un procesador y un sistema de almacenamiento de datos. Un programa informático es un conjunto de instrucciones que se pueden usar, directa o indirectamente, en un ordenador para llevar a cabo una determinada actividad o para lograr un determinado resultado.

- 5 Un programa informático puede ser escrito en cualquier forma de lenguaje de programación, incluyendo lenguajes compilados o interpretados, y puede ser desplegado en cualquier forma, incluyendo como un programa independiente o como un módulo, componente, subrutina, función, procedimiento u otra unidad adecuada para su uso en un entorno informático.
- 10 El programa informático puede almacenarse de manera no transitoria en un sistema de almacenamiento legible por ordenador. Ejemplos de sistemas de almacenamiento incluyen, sin limitación, discos ópticos tales como el CD, DVD y discos Blu-ray (BD); discos magnetoópticos; medios magnéticos tales como cintas magnéticas y discos duros internos y discos extraíbles; dispositivos de memoria semiconductores como EPROM, EEPROM y memoria flash; y RAM. Un sistema de almacenamiento legible por ordenador puede transformarse físicamente de tal manera que
- 15 contenga un programa informático. Un sistema de almacenamiento legible por ordenador que comprende instrucciones ejecutables por ordenador está configurado físicamente de tal manera que hace que un ordenador que interactúa con el sistema de almacenamiento lleve a cabo un proceso o un método.

- 20 Por tanto, en un aspecto, la divulgación proporciona un sistema de almacenamiento legible por ordenador que comprende instrucciones ejecutables por ordenador para llevar a cabo un método de alteración de una conformación de entrada de un polímero para generar una conformación alterada del polímero en un sistema informático, donde el polímero comprende (i) una pluralidad de residuos y (ii) una bisagra caracterizada por un parámetro de bisagra, comprendiendo el método (a) determinar la ubicación de la bisagra; y (b) aplicar una alteración al parámetro de bisagra, generando de este modo la conformación alterada de la proteína. En algunas realizaciones, un medio de
- 25 almacenamiento legible por ordenador comprende instrucciones ejecutables por ordenador que hacen que un ordenador lleve a cabo cualquiera de los métodos descritos en el presente documento.

- En otro aspecto, un sistema informático se divulga para llevar a cabo un método para alterar una conformación de entrada de un polímero para generar una conformación alterada del polímero, donde el polímero comprende (i) una
- 30 pluralidad de residuos y (ii) una bisagra caracterizada por un parámetro de bisagra, el sistema comprende un sistema de almacenamiento de datos y un procesador que comprende instrucciones para llevar a cabo un método que comprende (a) determinar la ubicación de la bisagra; y (b) aplicar una alteración al parámetro de bisagra. En algunas realizaciones, un sistema informático para llevar a cabo un método para alterar una conformación de entrada de un polímero para generar una conformación alterada del polímero, donde el polímero comprende (i) una pluralidad de
- 35 residuos y (ii) una bisagra caracterizada por un parámetro de bisagra, comprende un sistema de almacenamiento de datos y un procesador que comprende instrucciones para llevar a cabo cualquiera de los métodos descritos en el presente documento.

- 40 El procesador y el sistema de almacenamiento de datos pueden complementarse o incorporarse en circuitos integrados específicos de la aplicación (ASIC). Cuando se lee en el procesador del ordenador, que así se transforma físicamente, y se ejecutan o procesan posteriormente antes de la ejecución, las instrucciones del programa hacen que el ordenador programable lleve a cabo las diversas operaciones descritas anteriormente. El procesador y el sistema de almacenamiento de datos están comúnmente conectados por un bus.

- 45 Para prever la interacción con un usuario, la invención puede implementarse en un ordenador que tenga un dispositivo de visualización tal como, por ejemplo, un monitor de tubo de rayos catódicos (CRT) o pantalla de cristal líquido (LCD) para mostrar información al usuario. El usuario puede proporcionar entrada, por ejemplo, a través de un teclado y un dispositivo señalador, como un ratón. Las diversas conformaciones generadas por los presentes métodos pueden representarse gráficamente usando un software de modelado molecular y gráficos.

- 50 Los diferentes aspectos y realizaciones de la invención se pueden implementar en un sistema informático que incluye un componente de backend, tal como un servidor de datos, un componente de middleware, tal como un servidor de aplicaciones o un servidor de Internet, o un componente de interfaz, tal como un equipo cliente que tiene una interfaz de usuario, navegador de web o cualquier combinación de los mismos. Los componentes del sistema se pueden
- 55 conectar mediante cualquier forma o medio de comunicación de datos digitales.

- Los métodos actuales se pueden implementar en hardware en varias configuraciones. Por tanto, en algunas realizaciones, los procesos computacionales (tales como, por ejemplo, una pluralidad de simulaciones de dinámica molecular) se llevan a en paralelo en nodos de un clúster de ordenadores, en un sistema de computación distribuida
- 60 o en unidades de procesamiento de gráficos, como se entiende estas configuraciones en la técnica.

Aplicaciones

- 65 Los polímeros comúnmente experimentan cambios conformacionales a gran escala en un amplio intervalo de actividades funcionalmente críticas. Los métodos divulgados llenan un hueco importante para comprender mejor los subestados conformacionales de polímero y los factores energéticos que determinan las transiciones entre ellos. Esto,

a su vez, permite el diseño de polímeros optimizados para varios usos.

En una realización, el método de simulación de Monte Carlo/Dinámica Molecular descrito en el presente documento se usa para modelar, verificar y explicar los diversos estados conformacionales y los datos observados en los experimentos FRET (transferencia de energía de resonancia de fluorescencia) u otros métodos estructurales de baja resolución, como la microscopía crioelectrónica.

En un enfoque de diseño molecular guiado por estructura, ya sea en el contexto del diseño de fármacos o la ingeniería de proteínas, es deseable caracterizar la naturaleza dinámica de las proteínas y sus estados conformacionales alternativos. Por ejemplo, la geometría funcionalmente relevante de un polímero cuando se enlaza a su objetivo podría ser diferente de su geometría en el estado sin enlazar. El diseño de un fármaco o polímero puede optimizarse en función de estos estados conformacionales alternativos. En un aspecto, la presente divulgación proporciona un método para describir los estados de conformación alternativos de un polímero.

En una realización, el método se puede usar en el diseño y optimización de terapias basadas en proteínas con dos o más valencias, donde el término valencia, tal como se usa en el presente documento, se refiere al antígeno o componente de unión al objetivo del agente terapéutico. Estas moléculas terapéuticas multivalentes capaces de unirse a múltiples antígenos pueden verse impedidas estéricamente y estructuralmente a unirse a todos sus objetivos con todas sus valencias. En una realización, si el interés es optimizar o evaluar la capacidad de unión de un anticuerpo (por ejemplo, mostrado esquemáticamente en la figura 7) a su antígeno objetivo y coimplicar, por ejemplo, unir simultáneamente los antígenos objetivo con ambos brazos, los sistemas y métodos descritos en esta divulgación pueden ser empleados. Primero se modelaría la geometría de los dos objetivos entre sí. A continuación, se puede emplear el método descrito aquí para evaluar si la molécula candidata con la configuración de bisagra disponible es capaz de coimplicar ambos antígenos objetivo. En una realización alternativa, uno puede emplear los sistemas y métodos descritos para determinar la composición óptima de la bisagra que puede lograr la coimplicación. Tal enfoque también se puede emplear en el diseño y desarrollo de moléculas biespecíficas y otras moléculas multiespecíficas, multivalentes o multifuncionales, que implican más de una molécula simultáneamente. En algunas realizaciones, los sistemas y métodos se emplean para evaluar y modelar el acoplamiento de una molécula terapéutica a los receptores objetivo.

En un aspecto, la presente divulgación proporciona un método para producir un polímero en el que el polímero se diseña primero y luego se realiza. En realizaciones a modo de ejemplo, el polímero es una variante de polímero caracterizada por una propiedad mejorada de polímero en comparación con un polímero de referencia. La expresión "propiedades de polímero" se refiere a características físicas, químicas y biológicas que incluyen, entre otras, propiedades físicas (incluyendo el peso molecular, propiedades hidrodinámicas tales como el radio de rotación, carga neta, punto isoelectrónico, y propiedades espectrales como el coeficiente de extinción), propiedades estructurales (incluyendo las secundarias, terciarias, y elementos estructurales cuaternarios) estabilidad (incluyendo la estabilidad térmica, estabilidad en función del pH o de las condiciones de la solución, estabilidad de almacenamiento, y resistencia o susceptibilidad a la ubiquitinación, degradación proteolítica, o modificaciones químicas tales como la oxidación de metionina, deamidación de asparagina y glutamina, racemización de o epimerización de cadena lateral, e hidrólisis de enlaces peptídicos), solubilidad (incluyendo la susceptibilidad a la agregación en diversas condiciones, estado de oligomerización, y cristalizabilidad), propiedades cinéticas y dinámicas (incluyendo flexibilidad, rigidez, tasa de plegado, mecanismo de plegado, alosterismo, y la capacidad de sufrir cambios conformacionales y movimientos correlacionados), afinidad y especificidad de unión (a una o varias moléculas que incluyendo proteínas, ácidos nucleicos, polisacáridos, lípidos, y pequeñas moléculas, e incluyendo afinidades y tasas de asociación y disociación), actividad enzimática (incluyendo especificidad de sustrato; asociación, tasas de reacción y disociación; mecanismo de reacción; y perfil de pH), capacidad de modificación química o derivación (incluyendo la PEGilación y la fijación a otras moléculas o superficies), propiedades de expresión (como el rendimiento en uno o varios anfitriones de expresión, expresión corporal soluble frente a inclusión, localización subcelular, capacidad de ser secretado, y capacidad de mostrarse en la superficie de una celda), procesamiento y modificaciones postraslacionales (incluyendo el procesamiento proteolítico, glicosilación de N o C, lipidación, sulfatación, y fosforilación), propiedades farmacocinéticas y farmacodinámicas (incluyendo la biodisponibilidad después de la administración subcutánea, intramuscular oral o pulmonar; vida media sérica, distribución, y mecanismo y tasa de eliminación) y la capacidad de inducir un fenotipo alterado o una fisiología modificada (incluyendo la inmunogenicidad, toxicidad, capacidad de señalizar o inhibir señalización, capacidad para estimular o inhibir la proliferación celular, diferenciación, o migración, capacidad de inducir la apoptosis, y capacidad para tratar enfermedades).

Una vez diseñado, los polímeros se pueden fabricar fácilmente por métodos conocidos. Por ejemplo, en el caso de que el polímero sea una proteína, los métodos de expresión de proteínas utilizando ácido nucleico exógeno en células anfitrión pueden ser demandados. Tales enfoques son bien conocidos en la técnica y los detalles específicos de la implementación variarán con la célula anfitriona usada. El ácido nucleico exógeno, que codifica la proteína de interés, se puede realizar por síntesis génica total o por mutagénesis dirigida de un ácido nucleico que codifica proteína de tipo salvaje o variante. Métodos que incluyen ligadura dirigida por plantilla, PCR recursivo, mutagénesis de casete, se pueden utilizar mutagénesis dirigida al sitio u otras técnicas que son bien conocidas en la técnica.

El ácido nucleico exógeno puede formar parte de una construcción más grande, tal como un vector de expresión. En

la técnica se conocen numerosos tipos de vectores de expresión apropiados que comprenden el ácido nucleico exógeno y secuencias reguladoras adecuadas para varias células anfitrionas. Los vectores de expresión pueden contener secuencias reguladoras de la transcripción y la traslación seleccionadas de secuencias promotoras, sitios de unión ribosómicos, secuencias de inicio y detención transcripcionales, secuencias de inicio y detención de la traslacionales, señales de terminación de transcripción, señales de poliadenilación, secuencias potenciadoras o activadoras, genes marcadores seleccionables para permitir la selección de células anfitrionas transformadas, y similares.

Los vectores de expresión y el ácido nucleico exógeno que codifican la proteína de interés pueden introducirse en una célula huésped mediante técnicas que incluyen transfección mediada por dextrano, precipitación de fosfato de calcio, tratamiento con cloruro de calcio, transfección mediada por polibreno, fusión de protoplastos, electroporación, infección viral o de fago, encapsulación del (de los) polinucleótido(s) en liposomas, microinyección directa del ADN en núcleos y similares. Los ácidos nucleicos exógenos pueden integrarse de manera estable en el genoma de la célula huésped o pueden existir de forma transitoria o estable en el citoplasma.

Siguiendo la producción, los polímeros pueden aislarse o purificarse en varias formas conocidas por los expertos en la materia. Los métodos de purificación estándar incluyen técnicas cromatográficas, incluyendo intercambio de iones, interacción hidrofóbica, afinidad, dimensionamiento o filtración en gel, y fase inversa, llevada a cabo a presión atmosférica o a alta presión usando sistemas tales como FPLC y HPLC. Los métodos de purificación también incluyen técnicas electroforéticas, inmunológicas, precipitaciones, diálisis, y técnicas de cromatofocusing. Las técnicas de ultrafiltración y diafiltración, junto con los métodos de concentración de proteínas, también son útiles. Para orientación general en técnicas de purificación adecuadas, ver Scopes, Protein Purification: Principles and Practice (3rd ed., 1994), que se incorpora en el presente documento en su totalidad.

Las propiedades funcionales y biofísicas de los polímeros de la presente invención pueden seleccionarse en diversos ensayos, particularmente usando métodos in vitro. Los ensayos in vitro pueden permitir un amplio intervalo dinámico para seleccionar propiedades de proteínas de interés. Se pueden seleccionar múltiples propiedades simultáneamente o individualmente. Las moléculas pueden ser purificadas o no purificadas, dependiendo de los requerimientos del ensayo. Los ensayos de unión pueden llevarse a cabo usando varios métodos conocidos en la técnica, incluyendo, pero sin limitarse a, los ensayos basados en FRET (transferencia de energía de resonancia de fluorescencia) y BRET (transferencia de energía de resonancia de bioluminiscencia), AlphaScreen™ (Ensayo Homogéneo de Proximidad Luminiscente Amplificada), Ensayo de Proximidad de Centelleo, ELISA (Ensayo Inmunoabsorbente Ligado a Enzimas), SPR (Resonancia de Plasmón de Superficie, también conocido como BIACORE®), calorimetría de titulación isotérmica, calorimetría diferencial de barrido, electroforesis en gel, cromatografía que incluye filtración en gel y similares. Los ensayos pueden emplear varios métodos de detección usando, por ejemplo, etiquetas cromogénicas, fluorescentes, luminiscentes, o isotópicas.

Por consiguiente, la presente divulgación proporciona composiciones que comprenden una variante de polímero caracterizada por una propiedad física mejorada en comparación con un polímero de referencia. La variante de polímero puede diseñarse y fabricarse usando cualquiera de los métodos descritos en el presente documento. Las composiciones de la invención se pueden administrar a un paciente para tratar un trastorno. La administración de las composiciones puede tener lugar a través de varias rutas que incluyen rutas tópicas, enterales y parenterales.

La presente divulgación también proporciona composiciones farmacéuticas que comprenden una variante de polímero de la invención y un portador. Los portadores adecuados incluyen cargas tales como celulosa microcristalina, lactosa, com y otros almidones; agentes aglutinantes; edulcorantes y otros agentes saborizantes; agentes colorantes; polietilenglicol y similares. The Encyclopedia of Pharmaceutical Technology (James Swarbrick, ed., 3d ed. 2007), que se incorpora en el presente documento en su totalidad, proporciona un amplio intervalo de técnicas útiles para formular las proteínas de la presente invención.

Los artículos "un/a", "uno/a" y "el/la", tal como se usan en el presente documento no excluyen un número plural del referente, a menos que el contexto indique claramente lo contrario. La conjunción "o" no es mutuamente excluyente, a menos que el contexto indique claramente lo contrario. El término "incluir" se utiliza para referirse a ejemplos no exhaustivos.

Ejemplo 1

Se ha creado un prototipo que muestra muestras de regiones de bisagra de proteínas usando movimientos de Monte Carlo de ángulo torsional básico. Los movimientos se definen como los movimientos de ángulo torsional en las cadenas laterales y en la red troncal. El propósito de este método es muestrear el espacio disponible para dominios de proteínas grandes que están conectadas con una bisagra a otra parte del sistema que permanece fija, así como estimar el espacio conformacional disponible para carbohidratos ramificados en un sistema.

En este ejemplo, tal como se ilustra en la figura 8, una molécula basada en albúmina (albucore) se fusiona a dos scFvs unidos a cada terminal de la albúmina. Los enlazadores son A/334-A/350 así como B/255-B/270 y funcionan como una bisagra en esta molécula. Las regiones móviles son A/351 y superior, así como B/254 e inferior. Las cadenas

laterales razonables a mover son: A/352, A/378, A/547, A/547, A/450, A/293, A/290, A/294, A/333, A/297, B/314, B/371, B/307, B/311, B/225, B/224, B/150, B/149. Dado que el objetivo del método es estimar un límite superior en el volumen (espacio conformacional) muestreado por los dos scFv móviles enlazados por la bisagra a la albúmina, también se podría emplear un aumento de temperatura o un potencial de grano grueso para mejorar la eficacia de muestreo del algoritmo.

Usando el potencial AMBER y a temperatura de 10,000K, se ejecutaron cinco simulaciones de Monte Carlo que muestrearán los dos dominios scFv para 25.000 etapas a partir de la misma condición inicial (pero con diferentes semillas aleatorias). Las relaciones para los movimientos se establecieron en un 75 % de muestreo de torsión de red troncal y un 25 % de muestreo de torsión de cadena lateral. Con el fin de caracterizar las trayectorias y la orientación y la geometría de los dos scFv entre sí y los dominios de albúmina fijos, se calcularon las siguientes métricas para cada uno:

Uno. Se tomó una muestra de la distancia entre el residuo B/232 y A/395 (seleccionando un átomo en cada uno de estos residuos para calcular la distancia).

Dos. La distancia entre el residuo B/232 y la base de la bisagra, residuo B/271.

Tres. La distancia entre el residuo A/395 y la base de la bisagra, residuo A/333.

Cuatro. Usando la base de la bisagra para cada dominio como el origen de un sistema de coordenadas esféricas, se calcularon los ángulos polares y azimutales de la ubicación de los dos dominios a lo largo de la trayectoria.

Cinco. Usando residuo cerrado zapp, se estimaron los contactos entre los dominios móviles y el dominio fijo de albucore. Se realizó un gráfico que muestra el número de tramas donde cada residuo del dominio fijo está en contacto con los dominios móviles.

Seis. Para obtener una idea de la calidad de las conformaciones de la bisagra de la red troncal, se usó la cepa de red troncal zapp para verificar con qué frecuencia se han tensado las conformaciones de la red troncal en la región.

Siete. Los residuos que comprenden la bisagra se mutaron a diferentes composiciones de residuos para determinar el impacto en la flexibilidad. Como una mutación de prueba de todos los residuos de la bisagra a la prolina resultó en una bisagra menos flexible que la original, y este cambio se reflejó en el muestreo. En experimentos alternativos, todos los residuos de la bisagra se mutaron a alanina para aumentar las características de flexibilidad de la bisagra.

Mientras que AMBER es una función de energía basada en la física que comprende el valor de carga derivado del cálculo cuántico ab-initio y otros parámetros derivados empíricamente, se podrían llevar a cabo simulaciones usando una estructura de cristal de residuo/átomo de proteína más empírico derivada de potenciales estadísticos, también conocidos como potenciales basados en conocimiento.

La figura 18 es un ejemplo de un guion para el muestreo de bisagra zapp de acuerdo con este ejemplo.

La figura 9 muestra la distribución de ángulos aceptados para tres categorías de residuos de bisagra en este ejemplo: base de la bisagra, centro de la bisagra y uno más cercano al dominio en movimiento. Los ángulos phi y psi se trazan por separado. Los ángulos aceptados para las torsiones más cercanas a la base de la bisagra son significativamente más pequeños que los aceptados más cercanos a la región móvil. Esto se racionaliza sobre la base de que en ángulos de torsión más cercanos a la base de la bisagra resultarán en rotaciones del cuerpo fijas mucho más grandes.

En este ejemplo, la convergencia de muestreo del sistema albucore descrito anteriormente se caracterizó con las siguientes relaciones de aceptación AMBER: total 0,244, cadena lateral 0,747, enlazador 0,076, relaciones de aceptación DDRW: total 0,213, cadena lateral 0,849, enlazador 0,006 (para una ejecución de 10.000 etapas, 10.000 K). La figura 10 ilustra la distancia entre B/232 y A/395 en la simulación de Monte Carlo (5 ejecuciones, 25.000 etapas a 10.000 K).

Mutando los residuos de la bisagra a la prolina. La bisagra que reside en ambas cadenas se mutó a prolina usando una interfaz de programación de aplicaciones de mutaciones. Cuando el muestreo se realizó sobre la estructura resultante, las relaciones de aceptación para esta ejecución de muestreo fueron significativamente más bajas para la región de la bisagra (relación de aceptación: total 0,215, cadena lateral 0,742, enlazador 0,038). Cabe señalar que esto es aproximadamente la mitad de la relación de aceptación de la bisagra en comparación con la estructura de tipo salvaje, lo que implica que la bisagra es menos flexible. La figura 11 ilustra la distancia entre B/232 y A/395 en la simulación de Monte Carlo (5 ejecuciones, 25.000 etapas a 10.000 K) cuando los residuos de la bisagra se mutan a prolina.

Mutando todos los residuos de bisagra a la alanina. Por otra parte, la mutación de los residuos de la bisagra a la alanina aumentó significativamente la accesibilidad conformacional como resultado del movimiento de la bisagra, tal como se ilustra en la figura 12.

Resumen de Observaciones para el Sistema Albucore. Las distribuciones de distancia para las cinco ejecuciones de muestreo anteriores demuestran que las regiones móviles pueden acceder a los mismos volúmenes y no quedan atrapadas debido a diferentes semillas aleatorias. Cuando todos los residuos, excepto la región de bisagra, se mutan a alanina, la región móvil parece poder acceder a un volumen mayor, aunque esto podría deberse simplemente a un mayor número de movimientos aceptados. Por el contrario, cuando los residuos de la bisagra se mutaron a la prolina,

la relación de aceptación se redujo en un factor de dos, lo que implica que la bisagra es menos flexible que la de tipo salvaje en estas condiciones.

Las figuras 13 y 14 proporcionan capturas de pantalla de estructuras alternativas que representan conformaciones del dominio móvil.

5

Ejemplo 2

Se ha creado un prototipo que muestra muestras de regiones de bisagra de proteínas usando movimientos de Monte Carlo de ángulo torsional básico. En este ejemplo, se realiza muestreo de carbohidratos. Los movimientos se definen como los movimientos de ángulo torsional en el enlace glicosídico en el carbohidrato. El propósito de esta aplicación es muestrear el espacio conformacional disponible para grupos de carbohidratos ramificados que emplean el enlace glicosídico como una bisagra que conecta los azúcares y la proteína. El muestreo de carbohidratos se probó en la región del anticuerpo Fe en complejo con el Receptor FcR11a. Hay dos moléculas de carbohidratos en el sistema (cadena E, F) unidas a la porción de Fe del anticuerpo en la posición de residuos 297 en la cadena de anticuerpo en un residuo de Asn. La figura 15 proporciona una ilustración del complejo de Fe con las cadenas de carbohidratos E y F. Las dos cadenas de carbohidratos se tomaron muestras a 1000 K, 2000 K para más de 1.000.000 de etapas de Monte Carlo y se evaluaron las características ilustradas en la figura 19 para establecer su libertad de conformidad. Las figuras 17 y 18 son respectivamente dos histogramas para cada cadena de carbohidrato en la estructura de Fe.

Basado en esta ejecución, se hicieron las siguientes observaciones. Los gráficos de convergencia de los contactos de residuos muestran que las trayectorias comparadas muestran una superposición muy buena en términos de qué residuos están en contacto tanto para las simulaciones de Albucore como para el muestreo de carbohidratos. Parece que hay algunas excepciones donde ciertos residuos estuvieron en contacto en una trayectoria más a menudo que en otra, pero en general su número es pequeño y la frecuencia del contacto en la trayectoria no es tan alta. Las cifras que rastrean los movimientos aceptados de ciertos residuos en términos de ángulos de phi, psi frente a los resultados de estrés de la red troncal nos informan de que: (i) para la bisagra de tipo salvaje y los residuos mostrados, los ángulos phi-psi cambian con bastante frecuencia al muestrear conformaciones malas con mayor frecuencia que cuando muestrean bien (no hay problema de que simplemente se haga una transición a una conformación incorrecta y no se mueva en absoluto), y (ii) para la bisagra de la prolina y los residuos mostrados, las conformaciones están tensas a lo largo de la trayectoria, a partir de conformaciones erróneas (el hecho de que el valor phi permanezca constante tiene que ver con el hecho de que este ángulo implica un bucle cerrado y, por lo tanto, está excluido por TAS, por lo que las pequeñas diferencias mostradas en las figuras son numéricas).

REIVINDICACIONES

1. Un método de muestreo y análisis de la dinámica conformacional de polímeros mediante la búsqueda en el espacio de conformación de una proteína para determinar si una conformación tridimensional de la proteína puede unir más de un antígeno en una pluralidad de antígenos objetivo, comprendiendo la proteína una primera pluralidad de residuos, comprendiendo el método: en un sistema informático (11) que tiene uno o varios procesadores (22) y memoria (36) que almacena uno o varios programas a ejecutar por el uno o varios procesadores (22):
 - (A) obtener (402), desde el almacenamiento de memoria (36), un conjunto inicial de coordenadas tridimensionales $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$; (46) correspondientes a los átomos de la proteína, en donde la proteína comprende una pluralidad de dominios, cada respectivo X_{iA} en $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$ (46) corresponde a una coordenada tridimensional para un átomo en un primer dominio en la pluralidad de dominios, cada respectivo X_{iB} en $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$ (46) corresponde a una coordenada tridimensional para un átomo en un segundo dominio en la pluralidad de dominios, y cada respectivo X_{iC} en $\{X_{1A_init}, \dots, X_{NA_init}, X_{1B_init}, \dots, X_{MB_init}, \dots, X_{PC_init}, \dots\}$ (46) corresponde a una coordenada tridimensional para un átomo en una primera bisagra de la proteína, en donde la primera bisagra comprende una segunda pluralidad de residuos que es un subconjunto de la primera pluralidad de residuos, en donde la proteína está **caracterizada por** una capacidad para que el primer y el segundo dominios pivoten entre sí alrededor de la primera bisagra;
 - (B) alterar (404), utilizando un módulo de generación de polímeros (50), el conjunto inicial de coordenadas tridimensionales (46) de la proteína pivotando el primer dominio con respecto al segundo dominio sobre la primera bisagra, obteniendo así un conjunto alterado de coordenadas tridimensionales $\{X_{1A_alt}, \dots, X_{NA_alt}, X_{1B_alt}, \dots, X_{MB_alt}, \dots, X_{PC_alt}, \dots\}$ (56) para la proteína, en donde todos los átomos dentro del primer dominio se mantienen fijos entre sí durante la alteración, y todos los átomos dentro del segundo dominio se mantienen fijos entre sí durante la alteración;
 - (C) puntuar (406), usando un módulo de puntuación (52), una energía potencial calculada del conjunto alterado de coordenadas tridimensionales (56) frente a una energía potencial calculada de las coordenadas tridimensionales iniciales para la proteína con un criterio de Metropolis, en donde, cuando se cumple el criterio de Metropolis, el conjunto alterado de coordenadas tridimensionales es aceptado como el conjunto inicial de coordenadas tridimensionales;
 - (D) llevar a cabo (408), usando el módulo de puntuación (52), instancias adicionales de la alteración (B) y la puntuación (C) hasta que una energía del conjunto alterado de coordenadas tridimensionales $\{X_{1A_alt}, \dots, X_{NA_alt}, X_{1B_alt}, \dots, X_{MB_alt}, \dots, X_{PC_alt}, \dots\}$ (56) satisface el criterio de Metropolis; y
 - (E) evaluar si el conjunto alterado de coordenadas tridimensionales $\{X_{1A_alt}, \dots, X_{NA_alt}, X_{1B_alt}, \dots, X_{MB_alt}, \dots, X_{PC_alt}, \dots\}$ puede unir más de un antígeno en la pluralidad de antígenos de destino acoplando el conjunto alterado de coordenadas tridimensionales $\{X_{1A_alt}, \dots, X_{NA_alt}, X_{1B_alt}, \dots, X_{MB_alt}, \dots, X_{PC_alt}, \dots\}$ a un modelo de la pluralidad de antígenos.
2. El método de la reivindicación 1, comprendiendo el método, además, antes de la alteración (B): determinar, mediante implementación informática en el sistema informático (11), los residuos de la bisagra basándose en la accesibilidad del solvente y los contactos hechos por la segunda pluralidad de residuos en el conjunto inicial de coordenadas tridimensionales.
3. El método de la reivindicación 1, en donde la proteína comprende una pluralidad de bisagras, incluyendo la pluralidad de bisagras la primera bisagra, en donde, para cada bisagra respectiva en la pluralidad de bisagras, hay un par correspondiente de dominios en la pluralidad de dominios que pivotan entre sí en torno a la bisagra respectiva, comprendiendo el método, además, antes de la alteración (B): determinar, mediante implementación informática en el sistema informático (11), la identidad de los residuos en cada bisagra en la pluralidad de bisagras basándose en la accesibilidad del solvente y los contactos hechos por la segunda pluralidad de residuos en el conjunto inicial de coordenadas tridimensionales.
4. El método de la reivindicación 3, en donde la etapa de alteración (B) comprende, además, pivotar un tercer dominio con respecto a un cuarto dominio en la pluralidad de dominios alrededor de una segunda bisagra en la pluralidad de bisagras, en donde todos los átomos dentro del tercer dominio se mantienen fijos entre sí durante el giro del tercer dominio con respecto al cuarto dominio, y todos los átomos dentro del cuarto dominio se mantienen fijos entre sí durante el giro del tercer dominio con respecto al cuarto dominio.
5. El método de la reivindicación 1, en donde la primera bisagra comprende uno o varios parámetros de bisagra y la alteración (B) comprende (i) pivotar el conjunto inicial de coordenadas tridimensionales (46) de la proteína pivotando el primer dominio con respecto al segundo dominio sobre la primera bisagra sin aplicar un cambio a los uno o varios parámetros de la bisagra y (ii) reconstruir la primera bisagra en función de las ubicaciones del primer dominio y del

segundo dominio después del giro.

6. El método de la reivindicación 1, en donde la primera bisagra comprende una pluralidad de parámetros de bisagra y en donde menos de 2 ($n-2$) parámetros de bisagra se alteran en el giro del primer dominio con respecto al segundo dominio sobre la primera bisagra, en donde n es el número de residuos en la segunda pluralidad de residuos.

7. El método de la reivindicación 1, comprendiendo el método, además, antes de la alteración (B): determinar la ubicación de la primera bisagra sometiendo el conjunto inicial de coordenadas tridimensionales (46) al análisis de modo normal.

8. El método de la reivindicación 1, en donde la proteína comprende una pluralidad de bisagras, incluyendo la pluralidad de bisagras la primera bisagra, en donde, para cada bisagra respectiva en la pluralidad de bisagras, hay un par correspondiente de dominios en la pluralidad de dominios que pivotan entre sí en torno a la bisagra respectiva, comprendiendo el método, además, antes de la alteración (B) determinar la identidad de los residuos en cada bisagra en la pluralidad de bisagras

por análisis de componente principal o sometiendo el conjunto inicial de coordenadas tridimensionales (46) al análisis de modo normal, análisis de covarianza, análisis de dominio de proteína, análisis de rigidez, o de análisis de características de secuencia.

9. El método de la reivindicación 1, comprendiendo el método, además, antes de la alteración (B): determinar los residuos de la proteína que comprenden el primer dominio, la primera bisagra y el segundo dominio

por análisis de componente principal o sometiendo el conjunto inicial de coordenadas tridimensionales (46) al análisis de covarianza, análisis de dominio de proteína, análisis de rigidez, o análisis de características de secuencia.

10. El método de una cualquiera de las reivindicaciones 1-9, en donde la función de energía potencial comprende una función de energía basada en la física.

11. El método de una cualquiera de las reivindicaciones 1-9, en donde la función de energía potencial comprende una función de energía basada en el conocimiento.

12. El método de una cualquiera de las reivindicaciones 1-11, en donde el giro del primer dominio con respecto al segundo dominio comprende aplicar una alteración a uno o varios parámetros de bisagra de la primera bisagra, en donde cada parámetro de bisagra en uno o varios parámetros de bisagra comprende un ángulo de torsión de un residuo en la pluralidad de residuos de la primera bisagra.

13. El método de una cualquiera de las reivindicaciones 1-12, que comprende, además, realizar en el sistema informático (11) una simulación de dinámica molecular del conjunto alterado de coordenadas tridimensionales antes de la puntuación (C).

14. El método de una cualquiera de las reivindicaciones 1-13, en donde se realiza una pluralidad de instancias de la alteración (B) generando de este modo una pluralidad de conjuntos alterados de coordenadas tridimensionales (56).

15. El método de la reivindicación 1, en donde la proteína es un anticuerpo.

16. Un sistema informático (11) para buscar el espacio de conformación de una proteína para determinar una conformación tridimensional de la proteína que cumple una métrica de rendimiento, comprendiendo el sistema informático (11) al menos un procesador (22) y una memoria (36) que almacena al menos un programa para ser ejecutado por el al menos un procesador (22), comprendiendo la memoria (36), además, instrucciones para ejecutar el método de una cualquiera de las reivindicaciones 1-15.

17. Un medio de almacenamiento legible por ordenador no transitorio que almacena uno o varios módulos computacionales configurados para buscar el espacio de conformación de una proteína para determinar una conformación tridimensional de la proteína que cumple una métrica de rendimiento, comprendiendo el uno o varios módulos computacionales colectivamente instrucciones para llevar a cabo el método de una cualquiera de las reivindicaciones 1-16.

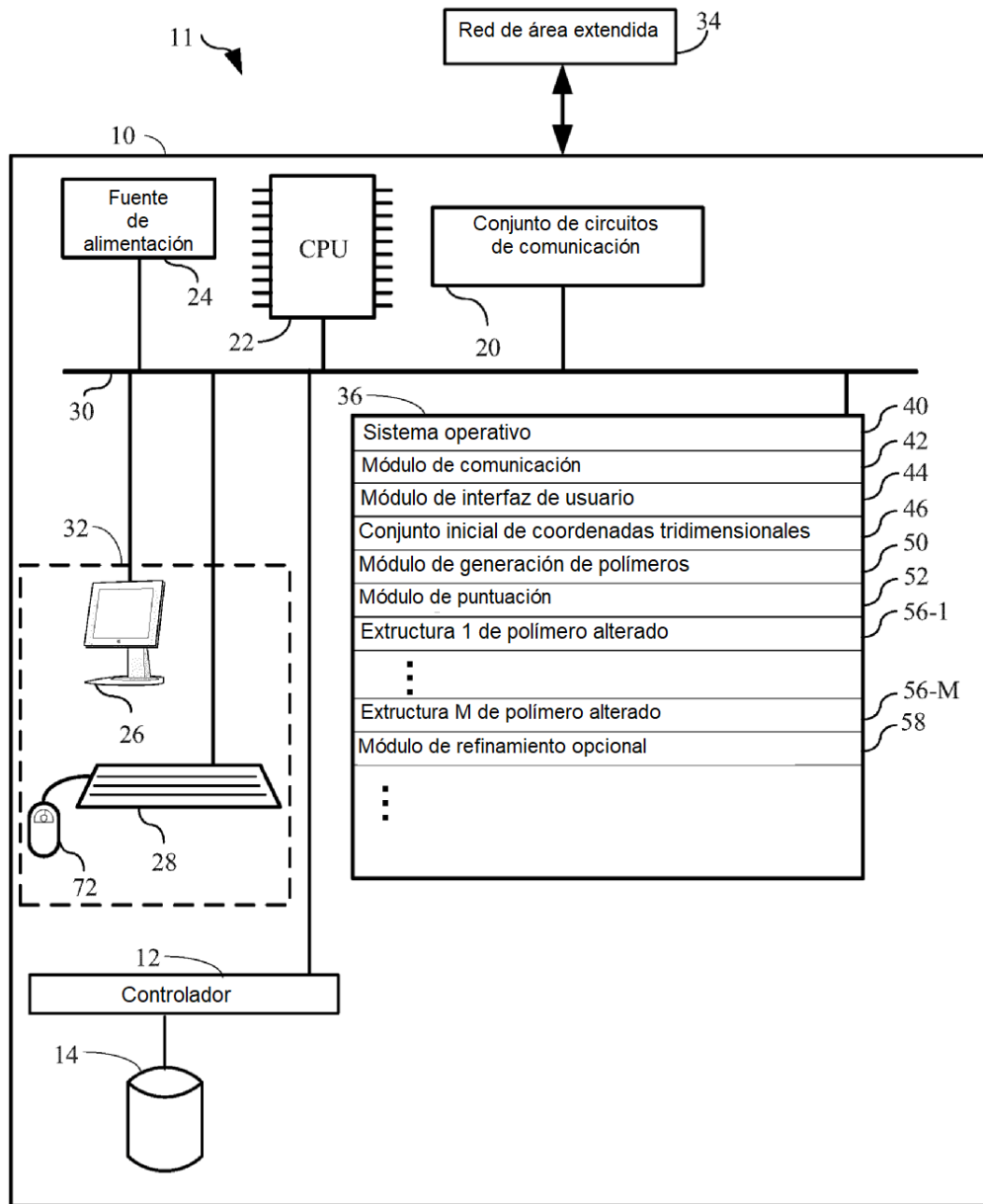


Fig. 1

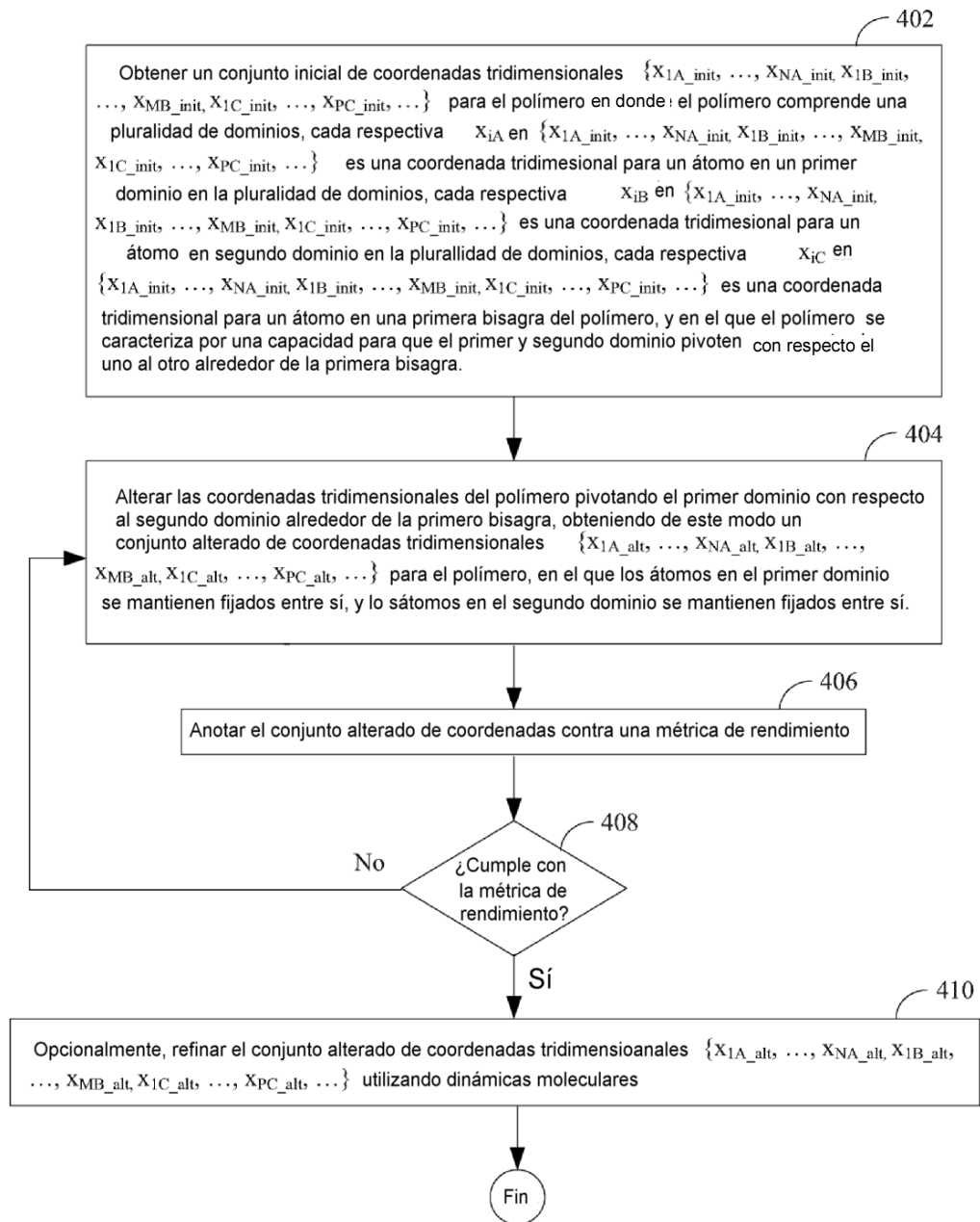


Fig. 2

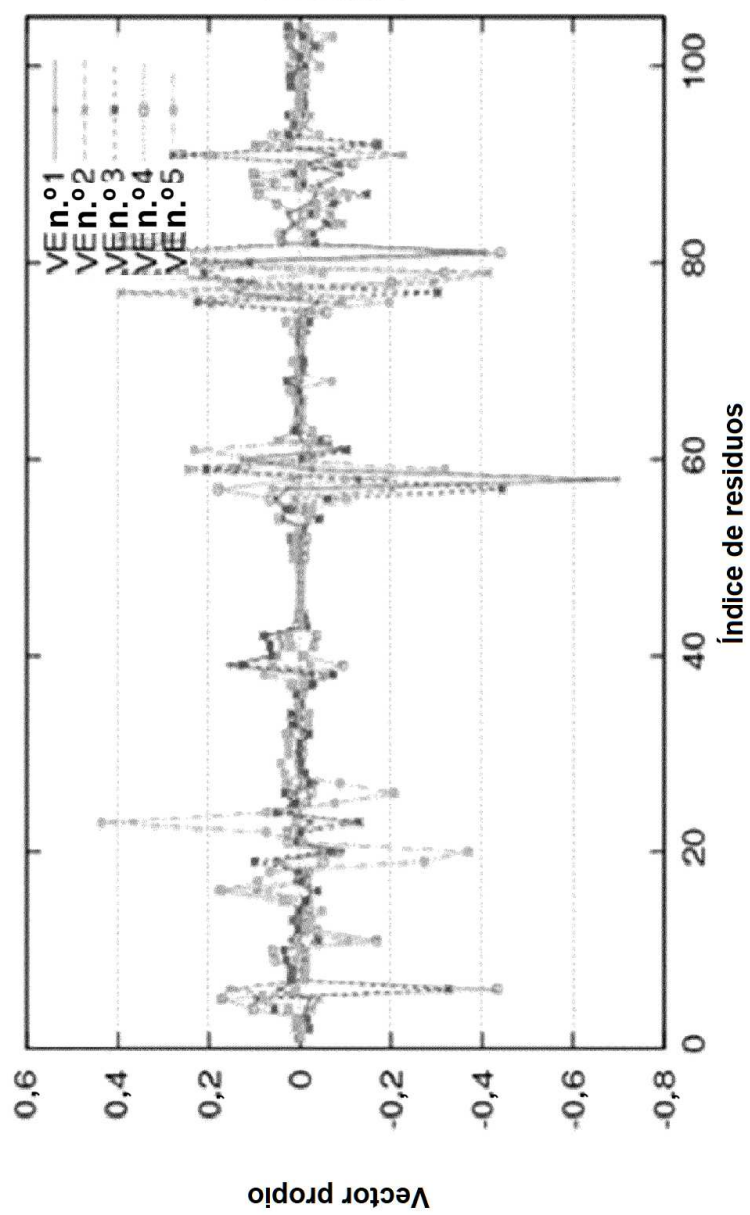
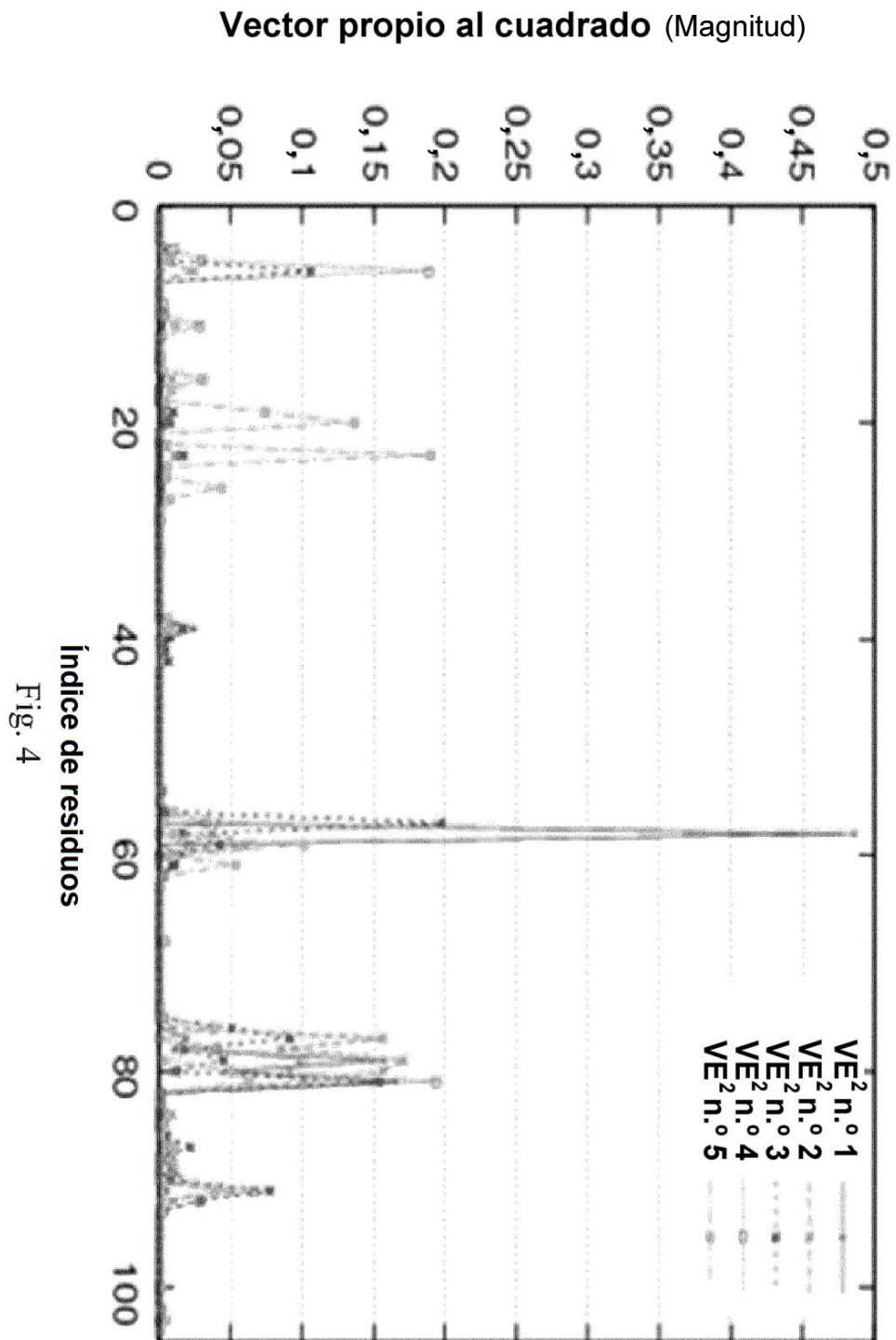


Fig. 3



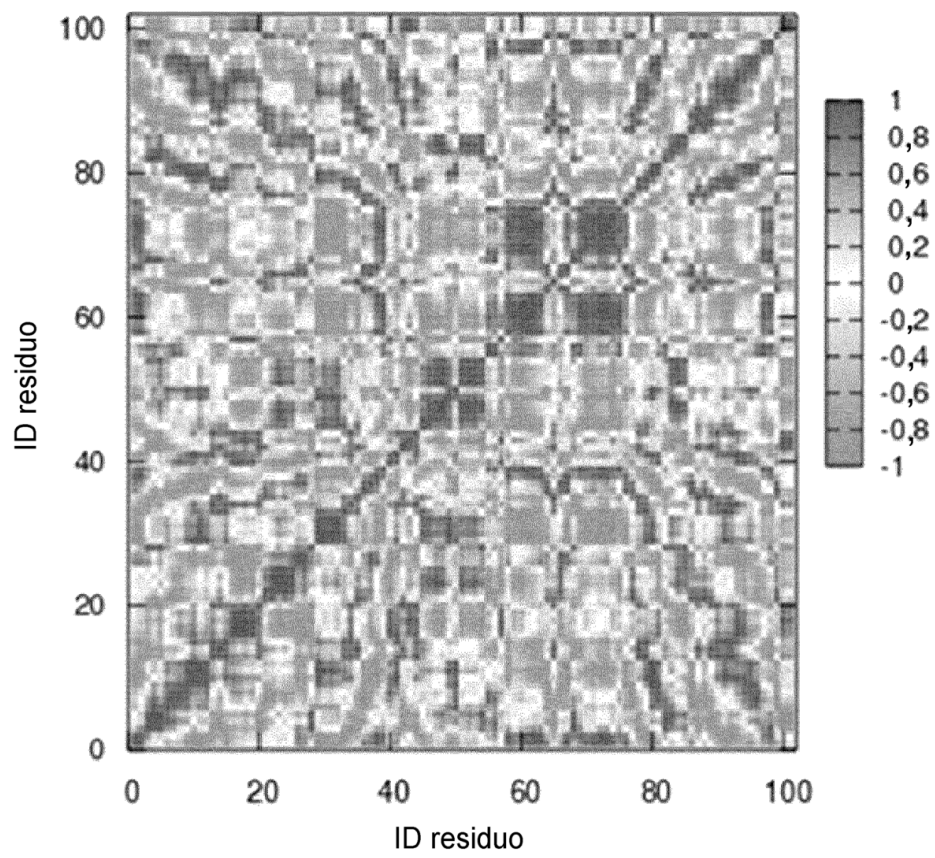


Fig. 5



Fig. 6

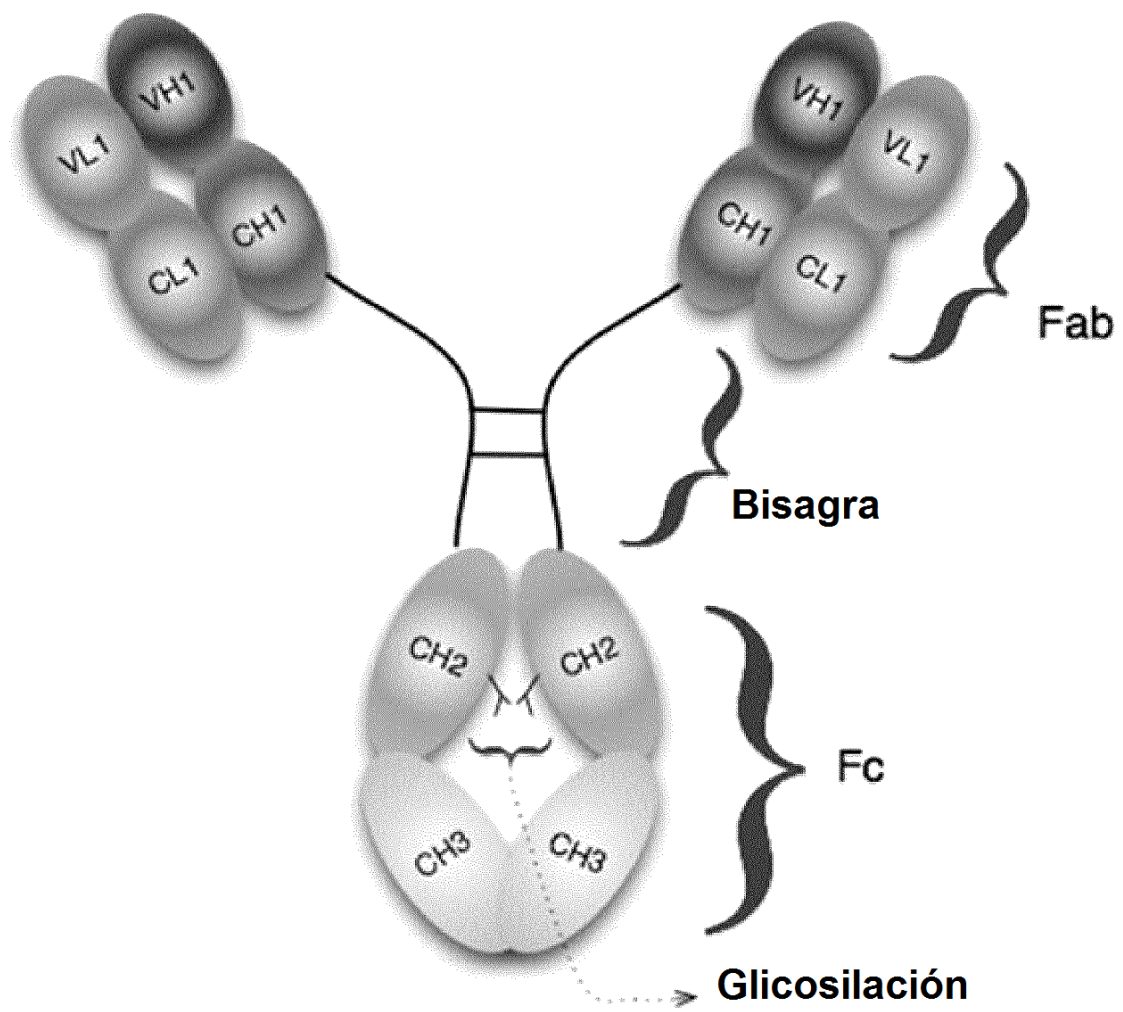


Fig. 7



Fig. 8

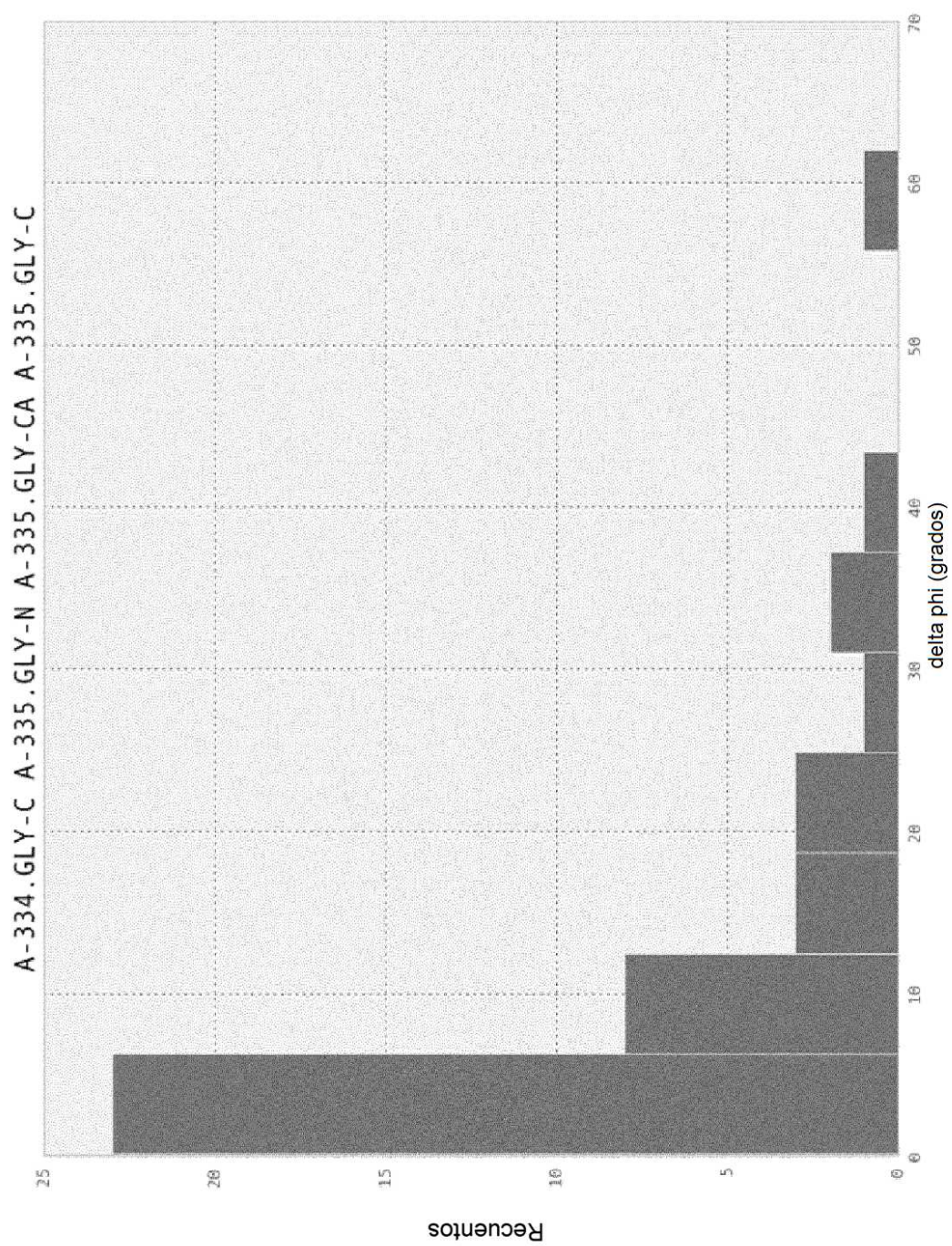
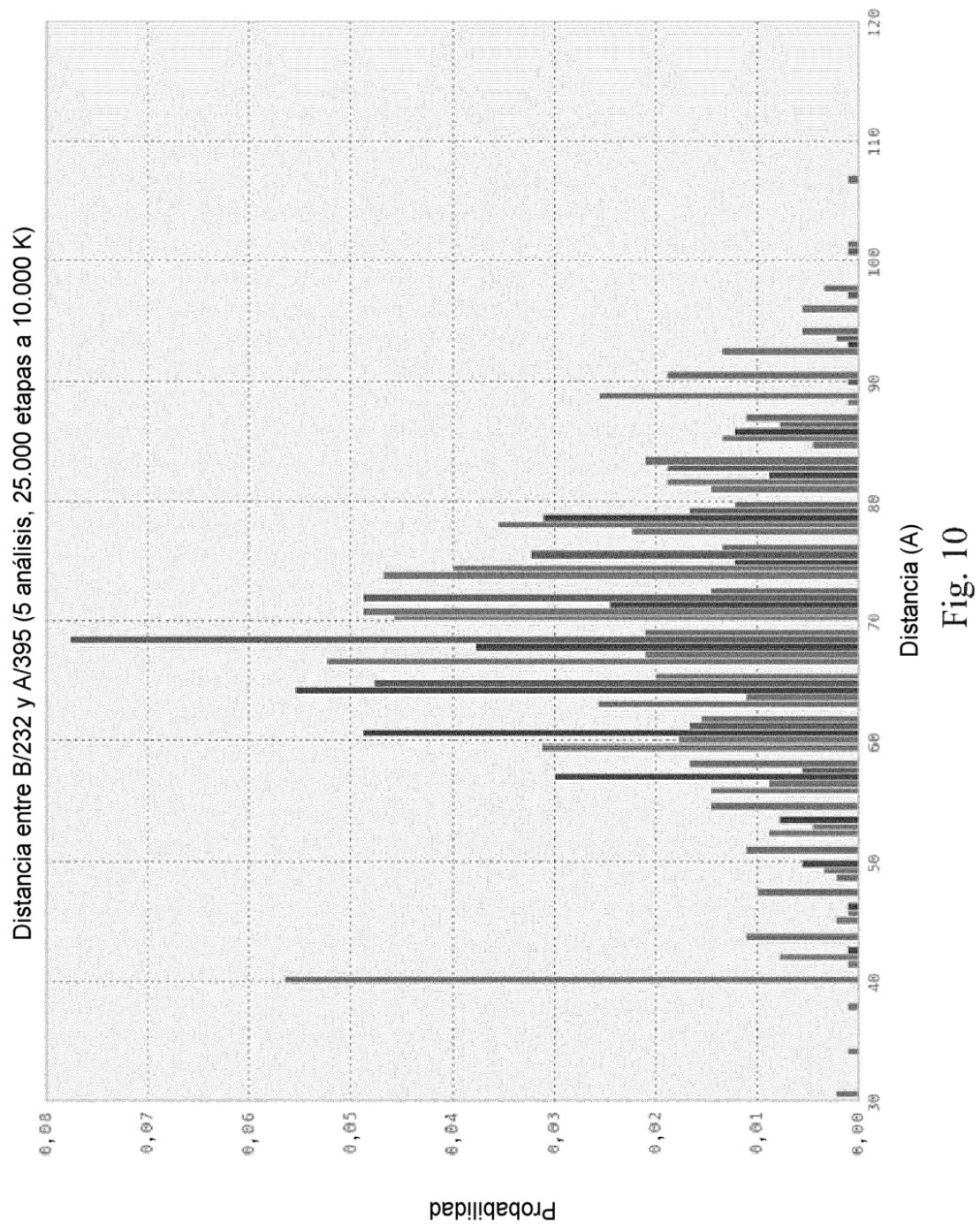


Fig. 9



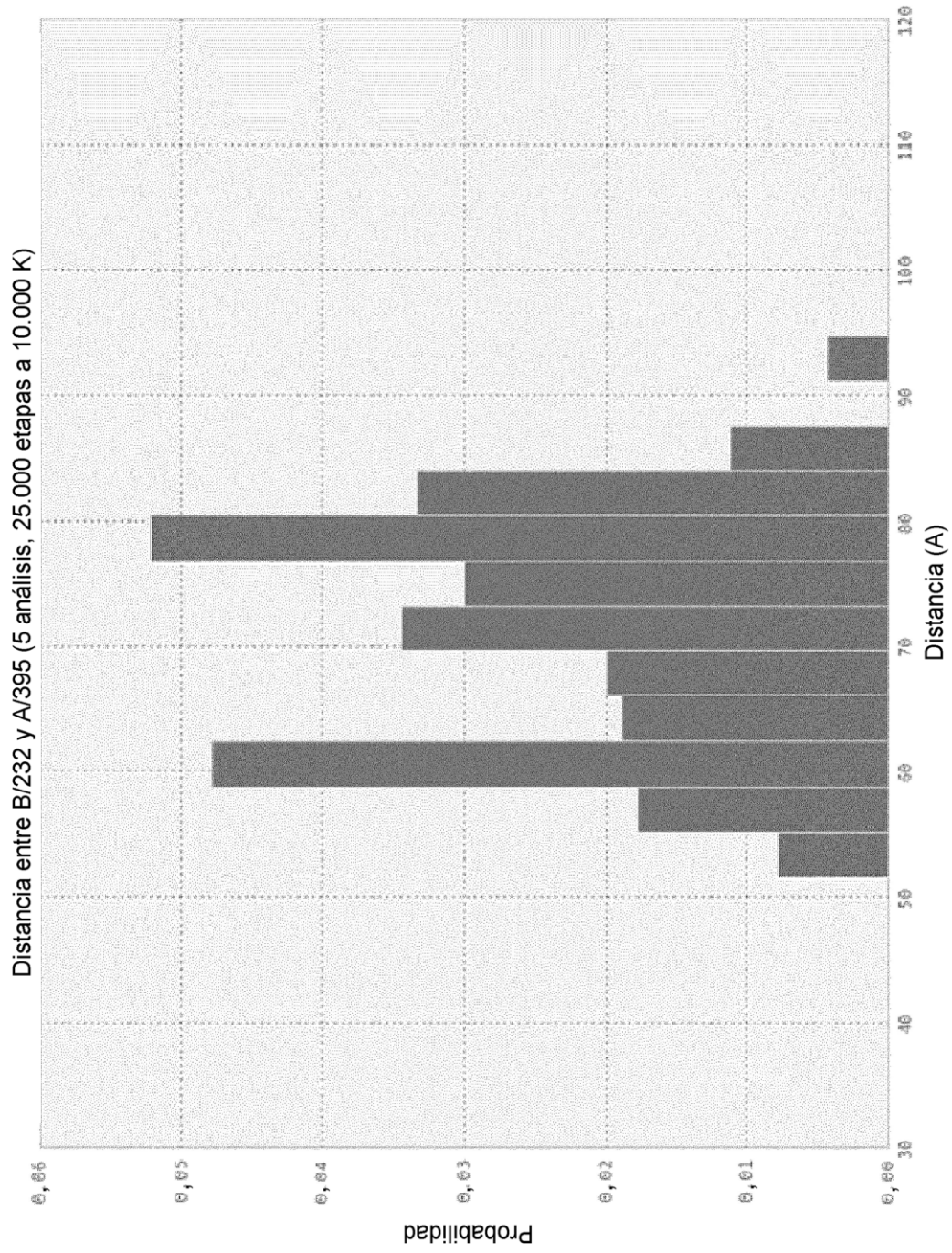


Fig. 11

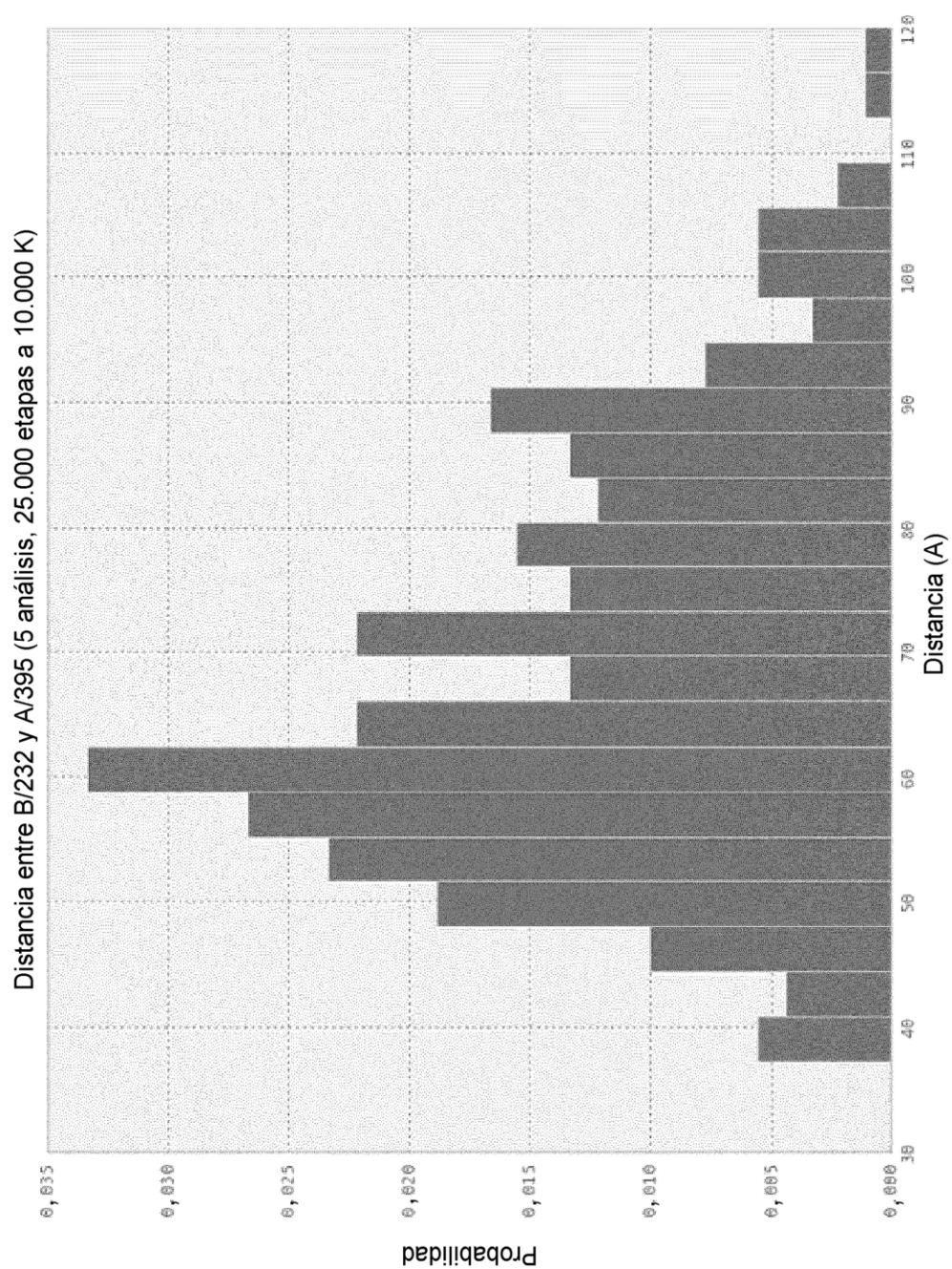


Fig. 12

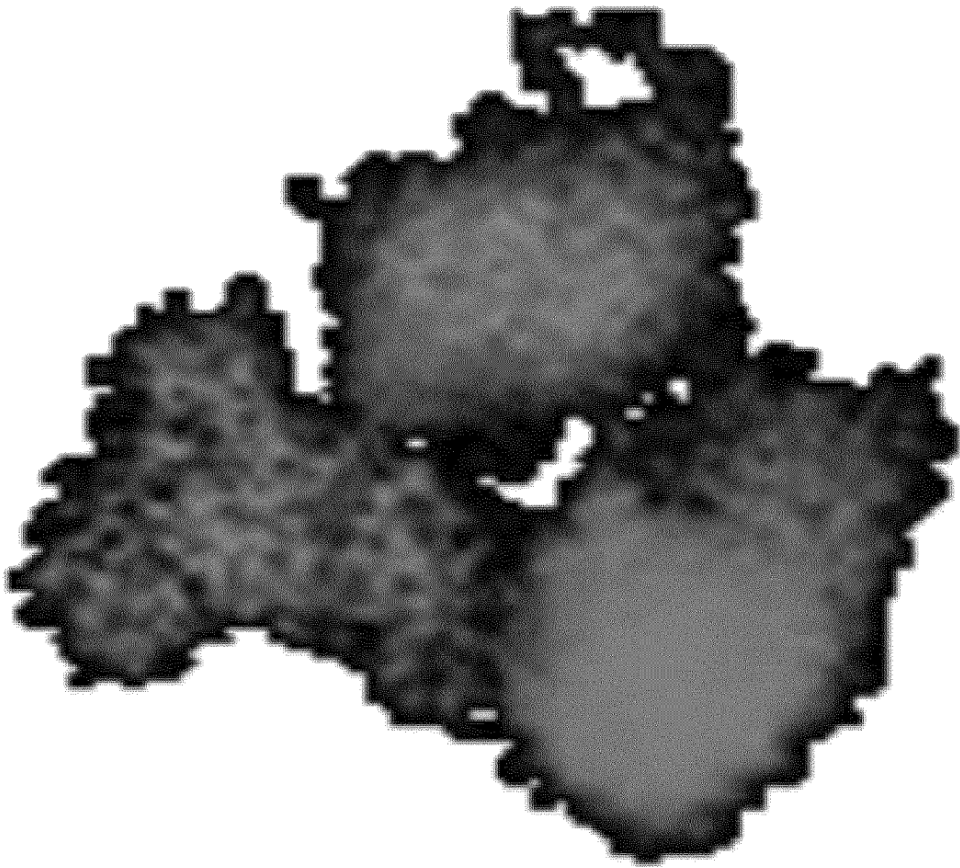


Fig. 13

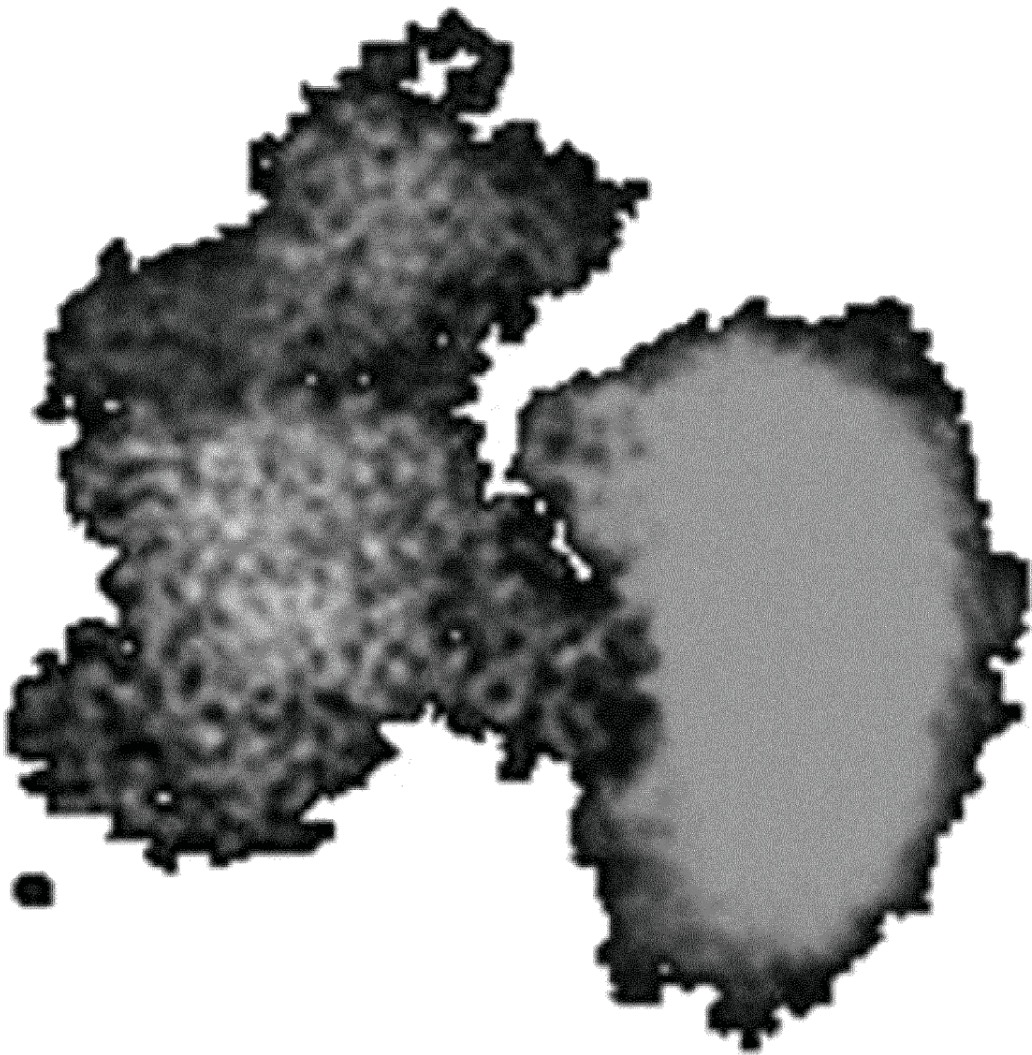


Fig. 14



Fig. 15

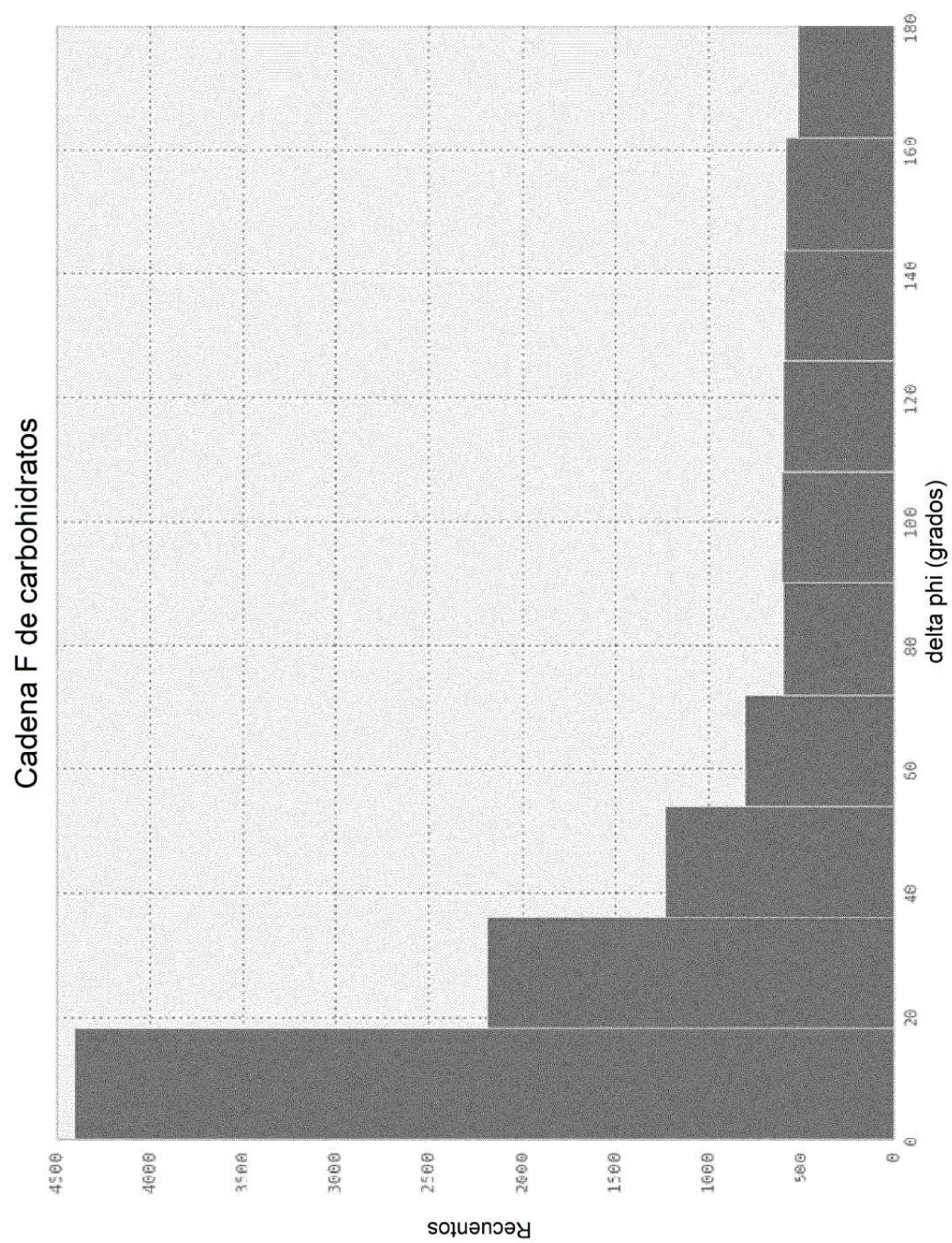


Fig. 16

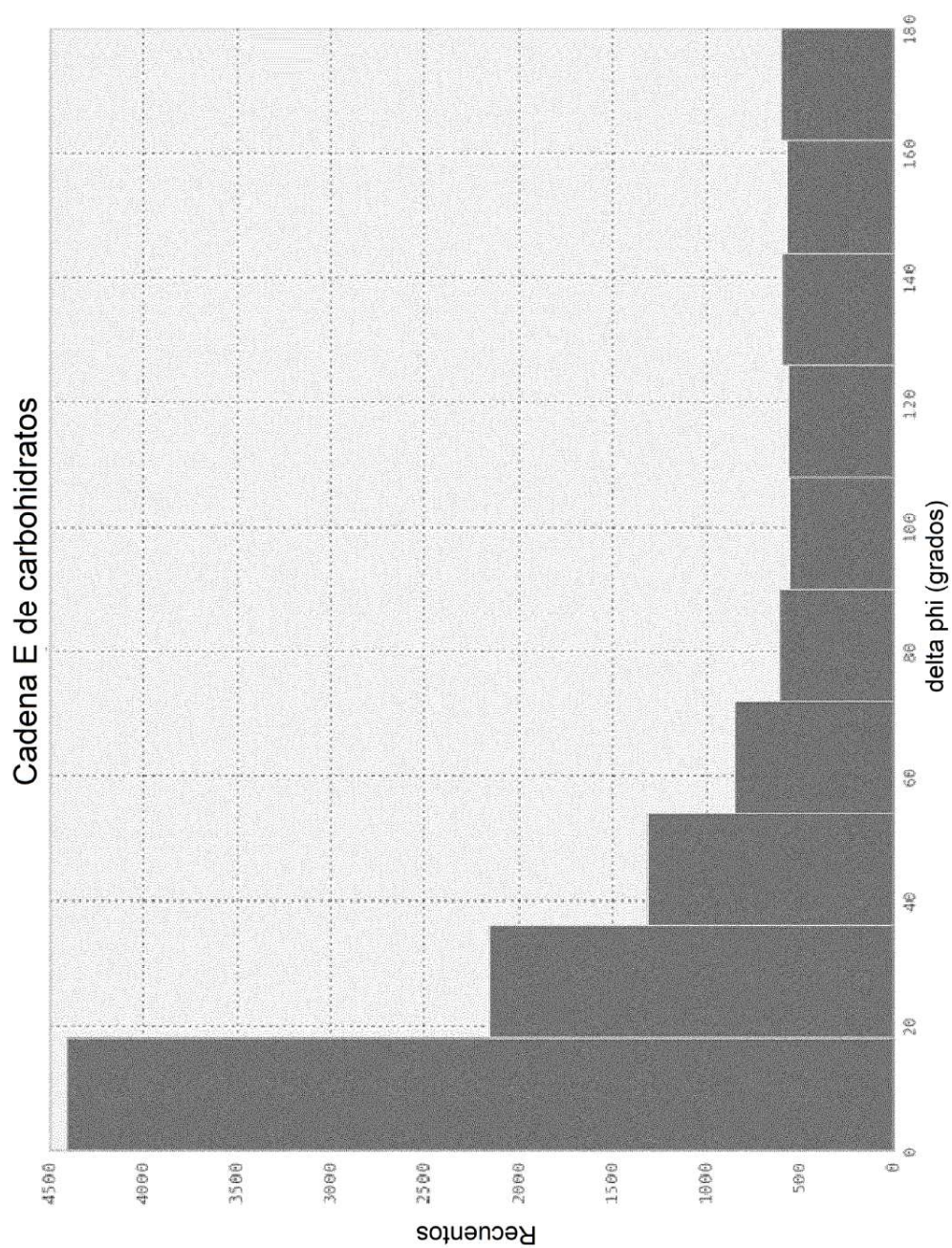


Fig. 17


```

zapp hinge_sampling --plib=$PLIB --rlib=$RLIB \
--struct=structures/
albul_nc_prime_bivalent_linked_4.azdb \
--sample-region="(residue.chain_id=A and
residue.seq_id >=334 and residue.seq_id <= 350)
or (residue.chain_id=B and residue.seq_id >= 255
and residue.seq_id <= 270)" \
--moving-region="(residue.chain_id=A and
residue.seq_id >=334 and residue.seq_id <= 350)
or (residue.chain_id=B and residue.seq_id >= 255
and residue.seq_id <= 270) or (residue.chain_id=A
and residue.seq_id >= 351) or (residue.chain_id=B
and residue.seq_id <= 254)" \
--sidechain-region="residue in [A/352, A/378, A/
547, A/547, A/450, A/293, A/290, A/294, A/333, A/
297, B/314, B/371, B/307, B/311, B/225, B/224, B/
150, B/149]" \
--steps=1000 \
--temperature=500 \
--snapshot-frequency=1 \
--out-structure=tas.azdb \
--out-trajectory=tas.nc \
--rand-seed=1 \
--debug

```

Fig. 18

```
zapp hinge_sampling --plib=$PLIB --  
rlib=$RLIB \  
--struct=../structures/IIIaF.azdb \  
--region-type="carbohydrate" \  
--sample-region="GLYCAM and not  
hydrogen" \  
--moving-region="GLYCAM" \  
--sidechain-  
region="residue.chain_id=A" \  
--steps=1000 \  
--temperature=20000 \  
--snapshot-frequency=1 \  
--sample-probability=1 \  
--sidechain-probability=0 \  
--out-structure=tas.azdb \  
--out-trajectory=tas.nc \  
--rand-seed=1 \  
--debug
```

Fig. 19