

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 808 724**

51 Int. Cl.:

**G01N 1/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **29.01.2015 PCT/US2015/013607**

87 Fecha y número de publicación internacional: **06.08.2015 WO15116862**

96 Fecha de presentación y número de la solicitud europea: **29.01.2015 E 15743662 (7)**

97 Fecha y número de publicación de la concesión europea: **08.04.2020 EP 3100023**

54 Título: **Métodos para muestrear soluto orgánico de entornos acuosos y heterogéneos**

30 Prioridad:

**29.01.2014 US 201461932890 P**

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

**01.03.2021**

73 Titular/es:

**UNIVERSITY OF MARYLAND, BALTIMORE  
(100.0%)  
Office of Technology Transfer, 620 W.Lexington  
Street, 4th Floor  
Baltimore MD 21201, US**

72 Inventor/es:

**MACKERELL, ALEXANDER, D., JR. ;  
RAMAN, E., PRABHU y  
LAKKARAJU, SIRISH, KAUSHIK**

74 Agente/Representante:

**FÚSTER OLAGUIBEL, Gustavo Nicolás**

**ES 2 808 724 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

Métodos para muestrear soluto orgánico de entornos acuosos y heterogéneos

5 **Declaración respecto a la investigación y el desarrollo con patrocinio federal**

Esta invención se realizó en parte con el respaldo del gobierno con la subvención número CA107331 otorgada por los Institutos Nacionales de Salud. El gobierno de los Estados Unidos tiene ciertos derechos en la invención.

10 **Campo de la invención**

La presente invención se refiere en general a métodos asistidos por ordenador para usar solutos orgánicos para muestrear entornos acuosos y heterogéneos. Más específicamente, se refiere a métodos asistidos por ordenador para determinar las distribuciones espaciales y la termodinámica de moléculas pequeñas en entornos acuosos y entornos heterogéneos, incluyendo entornos heterogéneos que contienen una proteína que tiene sitios de unión profundos u ocluidos. Los métodos proporcionados en el presente documento pueden usarse para estudios termodinámicos de sistemas acuosos complejos y para el diseño de fármacos asistido por ordenador.

20 **Antecedentes**

El potencial químico ( $\mu$ ) describe el movimiento de equilibrio de las partículas entre dos fases o estados. La fuerza impulsora detrás de este movimiento procede de 1) un gradiente de concentración, es decir, las partículas tienden a moverse desde una región de mayor concentración hasta una región de menor concentración para ganar entropía de mezcla; y 2) la afinidad química: las partículas se atraen a regiones de alta afinidad química (véase Dill KA & Bromberg S (2003), Chemistry and Biology).

El exceso de potencial químico ( $\mu_{ex}$ ) es el trabajo cuasiestático para llevar una partícula (por ejemplo, una molécula de soluto) desde la fase gaseosa hasta el disolvente;  $\mu_{ex} = \mu - \mu_{id}$ , donde  $\mu$  y  $\mu_{id}$  son el potencial químico y el potencial químico de gas ideal del soluto, respectivamente. En el contexto de la mecánica estadística, el potencial químico permite que el estado termodinámico de un sistema se defina en términos de un conjunto macrocanónico (GC) ( $\mu VT$ ) que permite la variación en las concentraciones de especies a través de fases/estados. Los procedimientos de simulación han evolucionado durante mucho tiempo hacia la determinación eficaz de la energía libre de hidratación de Gibbs (HFE), la afinidad química y otras propiedades termodinámicamente relevantes del agua y otras moléculas de soluto pequeñas a partir de conjuntos GC en lugar de los conjuntos isotérmicos, isobáricos (NPT), canónicos (NVT) o microcanónicos más convencionales donde se fija la concentración de la especie. Hasta la fecha, muchas de las estrategias del conjunto GC han empleado simulaciones de Montecarlo (MC) o bien para impulsar el muestreo de moléculas de agua o moléculas pequeñas individuales alrededor de proteínas o entornos cristalinos, o bien alternativamente para mejorar la precisión de los cálculos relativos de HFE en los cálculos de la perturbación de energía libre (FEP). Sin embargo, dado que las simulaciones de Montecarlo de conjunto macrocanónico (GCMC) de sistemas que contienen disolventes explícitos para representar la fase en volumen adolecen de problemas de convergencia debido a las bajas tasas de aceptación encontradas para la inserción de solutos, las simulaciones en el pasado se restringieron a la investigación de las afinidades químicas de solo el disolvente, las simulaciones que investigan las afinidades químicas para impulsar el muestreo de soluto individual en ausencia de disolvente explícito, que determinan las propiedades termodinámicas en condiciones cristalinas, o simulaciones para investigar el uso de estrategias de conjunto expandido.

Deng Y y Roux B (J Chem Phys. 21 de marzo de 2008; 128(11): 115103) describen una computación de energía libre de unión con dinámica molecular y simulaciones de Monte Carlo de conjunto macrocanónico. Clark M *et al.* (J Chem Inf Model. Abril de 2009; 49(4): 934-43) describen cálculos de energía libre de conjunto macrocanónico de unión de proteína-ligando.

En el contexto de entornos proteicos y macromoleculares, se han empleado técnicas de simulación de muestreo de fragmentos químicos para el descubrimiento o el diseño racional de moléculas que pueden unirse a dianas macromoleculares con altas afinidades para lograr un resultado biológico deseado. El método de identificación de sitio por saturación competitiva de ligando (SILCS) es una de tales técnicas que identifica la ubicación y las afinidades aproximadas de diferentes grupos funcionales en una superficie macromolecular diana mediante la realización de simulaciones de dinámica molecular (MD) de la diana en una disolución acuosa de moléculas de soluto que son representativas de diferentes fragmentos químicos. Sin embargo, estos conjuntos isotérmicos, isobáricos (NPT) de MD adolecen de las escalas de tiempo de difusión prolongadas de los solutos a través de entornos de macromoléculas y de disolvente explícitos, especialmente cuando los sitios de unión macromolecular están profundamente escondidos y son inaccesibles para el disolvente (es decir, ocluidos). Estas limitaciones implican que sólo pueden estudiarse los sitios de unión en una proteína que son accesibles para el disolvente en volumen. Sin embargo, una gran cantidad de proteínas biológicamente importantes, como los receptores acoplados a proteínas G (GPCR), las proteínas de receptores nucleares y otras macromoléculas, tienen cavidades de unión a ligandos profundas u ocluidas, y las simulaciones para estudiar la afinidad de las moléculas pequeñas por estas cavidades ocluidas generalmente requieren mucho tiempo.

Los métodos de la presente invención superan estas limitaciones, produciendo distribuciones espaciales precisas para solutos en entornos macromoleculares acuosos.

## 5 Sumario de la invención

La presente invención se define de la siguiente manera:

10 [1] Un método computacional para muestrear la distribución espacial de uno o más solutos y agua en una región del espacio definida (sistema) que comprende:

1) asignar una concentración objetivo,  $N_{obj}$ , de cada uno del uno o más solutos y agua;

15 2) muestrear la distribución espacial del uno o más solutos y agua en una región del espacio definida computacionalmente usando criterios de muestreo de Metropolis de Montecarlo de conjunto macrocanónico (GCMC), en el que el exceso de potencial químico ( $\mu_{ex}$ ) asignado para cada uno del uno o más solutos, si está presente, y agua se ajusta a 0, si está presente;

20 3) actualizar  $\mu_{ex}$  de cada uno del uno o más solutos y agua a partir de la diferencia en la concentración actual en la región del espacio definida ( $N_{sis}$ ) y el objetivo ( $N_{obj}$ ),

4) repetir las etapas 2) y 3) usando los valores actualizados de  $\mu_{ex}$  en la etapa 2) para obtener una distribución espacial del uno o más solutos y agua;

25 en el que el sistema comprende además una o más macromoléculas, en el que la distribución espacial de uno o más solutos o agua se usa para identificar la afinidad preferente de cada uno de los solutos o agua por cada una de la una o más macromoléculas, y en el que dicha una o más macromoléculas se seleccionan de una proteína, ARN, ADN, hidrato de carbono, lípido, compuesto químico orgánico, compuesto químico inorgánico o una combinación de los mismos.

30 [2] El método según [1], en el que el sistema contiene agua y un soluto y en el que la concentración objetivo del soluto y agua se ajusta a 1 M y 55 M, respectivamente, a partir de lo cual se obtiene la energía libre de hidratación del soluto a partir del valor de  $\mu_{ex}$

35 [3] El método según [1], en el que el  $\mu_{ex}$  de uno o más solutos y agua se aumenta y se disminuye alternativamente durante las operaciones de GCMC a través de ciclos consecutivos que implican las etapas 2)-4), una vez que la concentración de los solutos y el agua alcanzan su valor objetivo.

40 [4] El método según [1], en el que la proporción de etapas de GCMC para cada uno del uno o más solutos y agua se asigna basándose en la concentración objetivo de cada uno de los solutos y agua.

[5] El método según [1], en el que el sistema que contiene los solutos y/o agua está englobado en un sistema más grande que contiene dichos solutos y/o agua.

45 [6] El método según [5], en el que el sistema incluye adicionalmente una o más macromoléculas.

[7] El método según [1], en el que se muestrea la distribución espacial que sigue a la etapa 2) con una simulación de dinámica molecular.

50 [8] El método según [1], en el que  $\mu_{ex}$  se aumenta o se disminuye igual a  $N_{obj}/N_{sis}$ .

[9] El método según [1], en el que  $\mu_{ex}$  se aumenta cuando  $N_{sis}$  es menor que  $N_{obj}$  y se disminuye cuando  $N_{sis}$  es mayor que  $N_{obj}$ .

55 [10] El método según [1], en el que dicha región definida computacionalmente comprende una región interior que contiene el uno o más solutos y agua, ubicada dentro de una región exterior más grande que contiene agua adicional.

[11] El método según [1], en el que se genera una salida de las distribuciones espaciales del uno o más solutos y agua.

## 60 Breve descripción de dibujos

En el presente documento se describen realizaciones a modo de ejemplo no limitativas, con referencia a las siguientes figuras:

65

- La figura 1 ilustra una configuración de ejemplo para las simulaciones de GCMC-MD del estado convencional (esquema I) y la mezcla de soluto acuoso (esquema II). Las moléculas de agua y soluto se intercambian entre sus respectivos depósitos y los sistemas de simulación esféricos, A, indicados por el límite discontinuo, definido por el radio  $r_A$ . El sistema A está sumergido en un sistema más grande, B definido por el límite continuo, que incluía aguas adicionales de radio  $r_B$  ajustado a  $r_A + 5 \text{ \AA}$  en el presente estudio para impedir que los solutos hidrófobos ocupen el borde del sistema A (véase la figura 6). Alternativamente, el sistema más grande en el que está sumergido el sistema A puede tratarse usando condiciones de contorno periódicas e incluir otras entidades químicas además de agua (véase la figura 10).
- La figura 2 representa  $\mu_{\text{ex}}$  y concentraciones de los solutos y el agua en función de las iteraciones de GCMC-MD. Se varió  $\mu_{\text{ex}}$  para los solutos y el agua cada iteración basándose en sus concentraciones respectivas en el sistema de simulación A. La concentración objetivo de los solutos fue 1 M y 0,25 M en el esquema I (línea continua) y en el esquema II (línea discontinua) respectivamente, mientras que el agua se mantuvo a una concentración en volumen de 55 M en ambos sistemas. Los solutos son benceno (BENC), propano (PRPN), acetato (ACET), metilamonio (MAMM), metanol (MEOH), formamida (FORM) y acetaldehído (AALD).
- La figura 3 muestra: (A) Concentración; (B)  $\mu_{\text{ex}}$  de acetaldehído; y las probabilidades promedio de inserción (C) + deleción ( $P_{\text{ins+del}}$ ) y (D) traslación + rotación ( $P_{\text{tras+rot}}$ ) en función del número de iteraciones de GCMC-MD a partir de GCMC-MD de los esquemas I y II de acetaldehído con  $\mu_{\text{ex}}$  fijado en la HFE (cuadrados negros, línea discontinua) o fluctuado por  $d\mu_{\text{ex}}$  (línea continua, círculos en blanco) respectivamente. Obsérvese que el número de intercambios de soluto con el depósito de fase gaseosa es mayor con la fluctuación de  $\mu_{\text{ex}}$  mientras que la concentración y  $\mu_{\text{ex}}$  promedio son aproximadamente equivalentes.
- La figura 4 ilustra FragMaps de GFE seleccionados en el sitio de unión a ligando de la lisozima L99A de T4 a partir de una simulación de GCMC-MD de 10 x 50 ns y las conformaciones cristalinas minimizadas de los 9 ligandos (tal como se describe a continuación); se retiraron los átomos de proteína que ocuyen la vista de la cavidad de unión para una visualización clara. Los FragMaps se visualizan con un punto de corte de -1,2 kcal/mol para BENC (carbonos aromáticos de BENC) y PRPC (carbonos alifáticos de PRPN) y los FragMaps se visualizan con un punto de corte de -0,5 kcal/mol para AALO, MEOO, FORO (oxígenos aceptores polares de AALD, MEOH y FORH, respectivamente) y los grupos MEOH y FORH (hidrógenos donadores polares de MEOH y FORH, respectivamente).
- La figura 5 muestra la correlación de la afinidad de unión experimental  $\Delta G$  (de Morton A, Baase WA y Matthews BW (1995), Biochemistry 34 (27): 8564-8575) con las puntuaciones de LGFE para los nueve ligandos considerados (tal como se describe a continuación). Las puntuaciones de LGFE se obtienen de conjuntos conformacionales MD y MC de los ligandos (LGFE<sup>MD</sup>, LGFE<sup>MC</sup>) y los FragMaps de GFE. Los mapas generales tienen una correlación muy buena ( $R^2$  e índice predictivo, PI, altos).
- La figura 6 muestra la distribución de benceno en un sistema acuoso límite esférico. Cuando el soluto y las aguas comparten el mismo límite esférico, los solutos hidrófobos se acumulan en el borde de la pared. B. Para evitar esto, el sistema A se sumerge en un sistema B más grande tal como se describe en el texto principal.
- La figura 7 representa las funciones de distribución radial ( $g(r)$ ) para los átomos de soluto de las simulaciones de GCMC-MD y PBC MD de los esquemas I y II. Línea recta: MD con PBC que replican el esquema I, círculos en negro: MD con PBC que replican el esquema II, línea discontinua: GCMC-MD esquema I, círculos en blanco: GCMC-MD esquema II. Para benceno (BENC), propano (PRPN), metilamonio (MAMM) y acetato (ACET), se mide  $g(r)$  a través de las partículas sin masa (LP) que se añadieron a su centro de masas para impedir la agregación. Para metanol (MEOH), formamida (FORM) y acetaldehído (AALD), se mide  $g(r)$  entre los hidrógenos u oxígenos polares.
- La figura 8 muestra la concentración (M) y  $\mu_{\text{ex}}$  (kcal/mol), en función del número de ciclos de GCMC-MD del esquema I y el esquema II de los sistemas acuosos GCMC-MD con  $\mu_{\text{ex}}$  fijado en HFE (línea y círculos en negro) o fluctuado por  $d\mu_{\text{ex}}$  (cuadrados en negro y círculos en blanco) respectivamente.
- La figura 9 representa  $\mu_{\text{ex}}$  (kcal/mol), probabilidades de inserción + deleción ( $P_{\text{ins+del}}$ ) y traslación + rotación ( $P_{\text{tras+rot}}$ ) en función del número de ciclos de GCMC-MD del esquema I y GCMC-MD del esquema II de sistemas acuosos con  $\mu_{\text{ex}}$  fijado en HFE (línea continua y círculos en negro) o fluctuado por  $d\mu_{\text{ex}}$  (cuadrados en blanco y triángulos en blanco) respectivamente.
- La figura 10 ilustra una configuración de ejemplo para el GCMC-MD del esquema I con el mutante T4-L99A. El GCMC-MD está restringido a la esfera activa de radio de 20 Å (sistema A) con el centro en el sitio activo del mutante T4-L99A definido por los residuos Ala 99 y Met 102. El sistema A está englobado en el sistema B, que está una caja de PBC con paredes alejadas aproximadamente 12 Å de la superficie de la proteína. El sistema B contiene tanto agua como benceno a 55 M y 1 M, respectivamente. El agua y el benceno dentro de la esfera activa se indican mediante representaciones de líneas discontinuas y bolas y varillas, respectivamente.
- La figura 11 representa el coeficiente de solapamiento (OC) entre dos esferas de radio R, en función de la distancia

de separación entre ellas.

La figura 12 representa (A) FragMaps de GFE a partir de las simulaciones B2A (izquierda) y B2I (centro) solapadas en los estados activo (PDB: 3P0G) e inactivo (PDB: 2RH1) de los  $\beta_2$ AR con ligandos BI167107 y carazolol, respectivamente; se retiraron los átomos receptores que ocluyen la vista de la cavidad de unión. Los mapas diferenciales (derecha) resaltan las diferencias entre los dos estados. Los FragMaps de HBACC y HBDON se establecen en un punto de corte de -0,5 kcal/mol, mientras que los FragMaps de APOLAR, NEG y PDON se establecen en un punto de corte de -1,2 kcal/mol y el color para los FragMaps de compuesto no polar (APOLAR), donador neutro (HBDON), aceptor neutro (HBACC), aceptor negativo (NEG) y donador positivo (POS) es verde, azul, rojo, naranja y cian, respectivamente; (B) Las puntuaciones de LGFE se obtienen de conjuntos conformacionales de MC de agonistas, agonistas parciales (numerados 1-10), antagonistas y agonistas inversos (numerados 11-21) conocidos en los FragMaps tanto de B2A como B2I. LGFE y los valores experimentales de  $\Delta G_{unión}$  se correlacionan bien para agonistas y agonistas parciales con los FragMaps de B2A (1), y los antagonistas/agonistas inversos con los FragMaps de B2I (4); (C) Respuesta de relajación de anillos traqueales con los 7 primeros de los 15 ligandos seleccionados de los estudios de examen virtual (VS). Se usa isoproterenol (Iso, azul) como control.

La figura 13 representa FragMaps solapados en las LBP del AR (PDB 2AM9) con los ligandos A) TES, B) EM-5744 y C) S-1 en las orientaciones cristalográficas. D) FragMaps de la simulación sólo de GCMC. Se retiraron los átomos receptores que ocluyen la vista de la cavidad de unión para facilitar la visualización. El color para los FragMaps de compuesto no polar (APOLAR), donador neutro (HBDON), aceptor neutro (HBACC), aceptor negativo (NEG) y donador positivo (POS) es verde, azul, rojo, naranja y cian, respectivamente. Los FragMaps de APOLAR, HBACC y HBDON se establecen en un punto de corte de -0,5 kcal/mol, mientras que los FragMaps de NEG y POS se establecen en -1,2 kcal/mol. Las afinidades distintivas de FragMap que se solapan con los grupos funcionales de los ligandos se indican mediante flechas de color igual que los FragMaps. D) La ausencia de flexibilidad proteica en las simulaciones sólo de GCMC conduce a una disminución general en la extensión espacial de los FragMaps y a la omisión de FragMap de APOLAR A2 que está en las inmediaciones de las conformaciones cristalográficas de los segundos anillos de fenilo de B) EM-5744 y C) S-1. El muestreo de MC de ligandos, EM-5744 y S-1 produce conformaciones (amarillas) distintas del cristal (cian), en B y C.

La figura 14 representa FragMaps de PPAR $\gamma$  solapados en las LBP de PPAR $\gamma$  (PDB 3U9Q) con los ligandos A) ácido decanoico, B) rosiglitazona (PDB: 2PRG,) C) GW409544 (PDB: 1K74) y D) Cerco-A (PDB: 3B1M) en sus orientaciones cristalográficas; se retiraron los átomos receptores que ocluyen la vista de la cavidad de unión para facilitar la visualización. Los FragMaps de HBACC y HBDON se establecen en un punto de corte de -0,5 kcal/mol, mientras que los FragMaps de APOLAR, NEG y PDON se establecen en un punto de corte de -1,2 kcal/mol. No se encontraron FragMaps que se solaparan con la dibenzofurancaboxamida de Cerco-A (D).

La figura 15 representa que los FragMaps trazan el posible paso de entrada de un ligando en el PPAR $\gamma$ . El ligando GW40944 se muestra en las LBP.

La figura 16 representa FragMaps de GFE solapados en las LBP parcialmente ocluidas de mGluR1 con los ligandos A) FITM y B) MPEP. Se retiraron los átomos de proteína que ocluyen la vista de la cavidad para facilitar la visualización. Todos los contornos de FragMap se muestran a -1,2 kcal/mol y el color para FragMaps de compuesto no polar (APOLAR), donador neutro (HBDON), aceptor neutro (HBACC), aceptor negativo (NEG) y donador positivo (POS) es verde, azul, rojo, naranja y cian, respectivamente.

La figura 17 compara FragMaps entre A) GCMC/MD frente a B) Simulaciones sólo de GCMC para mGluR. Todos los contornos de FragMap se visualizan a -1,2 kcal/mol. C) LGFE calculadas usando los FragMaps a partir de GCMC/MD correlacionados bien con  $\Delta G_{unión}$  mientras que la correlación se perdió cuando D) las LGFE se calcularon utilizando los FragMaps a partir de simulaciones sólo de GCMC.

La figura 18 representa el análisis estructural de B2A y B2I. A) Distribuciones de RMSD de cadenas laterales seleccionadas identificadas como pertinentes para la unión del ligando a través de la simulación de GCMC/MD de 10x50 ns de B2A (naranja) y B2I (verde). Los residuos polares son de color verde, los residuos hidrófobos son de color morado y la Asp cargada de manera negativa es de color rojo. BI-167107 y el carazolol son de color naranja y verde respectivamente. B) Las conformaciones inactiva (verde) y activa (naranja) de  $\beta_2$ AR se mantienen a través de las simulaciones como lo demuestra la distribución del ángulo entre las hélices TM H5 y H6 y la distancia entre los átomos CE y CD a través de los residuos de formación del puente salino R131-E268.

La figura 19 representa los modos de unión de algunos agonistas y agonistas parciales en la conformación activada de B2A a partir del muestreo de MC de los ligandos en el campo de los FragMaps de B2A SILCS-GCMC/MD.

La figura 20 representa los modos de unión de algunos antagonistas y agonistas inversos en la conformación inactivada de  $\beta_2$ AR a partir del muestreo de MC de los ligandos en el campo de los FragMaps de B2I SILCS-GCMC/MD.

La figura 21 representa las estructuras de los quince ligandos seleccionados para estudios de evaluación funcional a partir del examen virtual impulsado por las diferencias en FragMaps entre los estados activo e inactivo de  $\beta_2$ AR.

5 La figura 22 representa las conformaciones acopladas de los ligandos preseleccionados L1, L3, L4 y L7 que se solapan bien con los diferentes FragMaps a partir de simulaciones B2A, lo que lleva a buenas puntuaciones de LGFE.

### Descripción detallada

10 La invención es tal como se define en las reivindicaciones adjuntas. Se proporcionan métodos asistidos por ordenador para usar solutos orgánicos para muestrear entornos acuosos y heterogéneos. En la siguiente descripción, con fines explicativos, se exponen numerosos detalles específicos para proporcionar una comprensión completa de la presente invención. Sin embargo, resultará evidente para un experto en la técnica que la presente invención puede ponerse en práctica sin estos detalles específicos. Todavía otros aspectos, características y ventajas de la invención son fácilmente evidentes a partir de la siguiente descripción detallada, simplemente ilustrando varias realizaciones e implementaciones particulares, incluyendo el mejor modo contemplado para llevar a cabo la invención. La invención también es capaz de otras y diferentes realizaciones, y sus diversos detalles pueden modificarse en varios aspectos obvios, siempre que sean según las reivindicaciones adjuntas. Por consiguiente, los dibujos y la descripción deben considerarse de naturaleza ilustrativa y no restrictiva.

20 Los métodos de la presente invención y los sistemas según ellos permiten la investigación del exceso de potencial químico ( $\mu_{ex}$ ) de solutos en disolución acuosa, incluyendo disoluciones que contienen múltiples solutos. Fundamental para el enfoque es el uso de un  $\mu_{ex}$  fluctuante a lo largo de las partes de GCMC de las simulaciones. Esto conduce a la convergencia de  $\mu_{ex}$  para el/los soluto(s) dado(s) y el entorno, según lo definido por el usuario, basándose en la concentración objetivo y en el mantenimiento del muestreo de soluto una vez que el sistema ha convergido con respecto a  $\mu_{ex}$  o la concentración objetivo. Además del examen de  $\mu_{ex}$  requerido para mantener los solutos a su concentración objetivo en ambientes acuosos, el GCMC-MD iterativo con enfoque de  $\mu_{ex}$  fluctuante también es útil para el muestreo eficaz de solutos.

30 Está bien aceptado que dirigir ligandos a cavidades de unión de ligando (LBP) ocluidas en proteínas con accesibilidad mínima o nula al entorno circundante es un desafío. Como se sabe que las eficacias de los ligandos tanto de los GPCR como de los receptores nucleares (NR) están asociadas a pequeños cambios conformacionales en sus sitios de unión, el modelado preciso de estos sitios es crítico para el desarrollo futuro de agentes terapéuticos para una amplia variedad de enfermedades.

35 Puede usarse la metodología de SILCS para mapear los patrones de afinidad de energía libre de grupos funcionales en las superficies de proteínas, incluyendo las LBP. Este método explica la flexibilidad conformacional de las proteínas, el espacio químico de los ligandos y el disolvente explícito mediante la realización de la dinámica molecular (MD) de la proteína objetivo en una disolución acuosa de pequeñas moléculas de soluto representativas de diferentes grupos químicos funcionales. En una realización, los patrones de afinidad de estos grupos funcionales se obtienen en forma de probabilidad discretizada, o mapas de energía libre, llamados FragMaps. La inclusión de la flexibilidad de la proteína y la representación del disolvente explícito en el método de SILCS es particularmente importante dados los cambios conformacionales conocidos dentro de la cavidad de unión tras la unión del ligando y la competencia y el desplazamiento del agua por los ligandos.

45 Tal como se describe a continuación, el método de SILCS fue satisfactorio en el mapeo de los requisitos de grupos funcionales de ligandos para una variedad de macromoléculas. Para examinar las LBP ocluidas, se combina SILCS con una metodología iterativa de Montecarlo de conjunto macrocanónico (GCMC) y MD. GCMC impulsa el muestreo de solutos pequeños y disolventes explícitos en las LBP, mientras que MD permite el muestreo conformacional de las macromoléculas en presencia de solutos y agua, lo cual es útil para explorar cavidades críticas ausentes en estructuras apocristalinas que se sabe que sirven como sitios de unión.

55 La presente invención describe el uso de SILCS-GCMC/MD para mapear patrones de afinidad de grupos funcionales de las cavidades ocluidas de las siguientes proteínas NR terapéuticamente importantes y GPCR. Macromoléculas terapéuticamente importantes a modo de ejemplo son el receptor de andrógenos (AR), el receptor  $\gamma$  activado por proliferador de peroxisomas (PPAR $\gamma$ ), los receptores nucleares y los GPCR, tal como el receptor de glutamato metabotrópico (mGluR) y el receptor  $\beta_2$ -adrenérgico ( $\beta_2$ AR). El método se utilizó para predecir las afinidades de unión relativas de ligandos usando un esquema de puntuación de energía libre en la cuadrícula de ligandos (LGFE) que se describe adicionalmente a continuación, e incorpora parámetros de flexibilidad conformacional de proteínas. El método también es capaz de distinguir entre los estados activo e inactivo del  $\beta_2$ AR a través de diferencias en los patrones de afinidad de ligandos en estos estados. Desde el punto de vista del descubrimiento de fármacos, tal información es útil para distinguir la función de los ligandos. Validar esta capacidad permite utilizar las diferencias de FragMaps para identificar nuevos agonistas de  $\beta_2$ AR, compuestos que tienen el potencial de convertirse en agentes terapéuticos para el tratamiento del asma y otras enfermedades pulmonares obstructivas.

65 En una realización, el método puede usarse para muestrear el espacio de configuración de una cavidad ocluida en

una macromolécula. Este muestreo puede realizarse en un solo soluto conjuntamente con agua, en el que la concentración del soluto *in silico* está en un intervalo de aproximadamente 0,01 M a 1000 M para estimar la afinidad de unión del soluto o una mezcla compleja de solutos por la cavidad ocluida de una macromolécula y para mapear la preferencia funcional de los grupos por la cavidad ocluida (de unión) de la macromolécula. En determinados ejemplos, la concentración de soluto es de 1 M porque corresponde a la concentración en el estado convencional, lo que proporciona la energía de unión libre experimental. Cuando esto se analiza en el contexto de la metodología de SILCS, se muestra que el enfoque reproduce cualitativamente la orientación de unión de ligandos conocidos, así como clasifica cuantitativamente el orden de las afinidades de unión de ligandos. Dicha información sobre la funcionalidad y la orientación de unión de los ligandos ayuda con los métodos de diseño de fármacos asistidos por ordenador que se dirigen a la identificación de ligandos de alta afinidad para macromoléculas con cavidades de unión profundas u ocluidas, tales como factores nucleares proteicos y GPCR.

La presente invención incluye, por tanto, métodos para el muestreo de solutos en sistemas acuosos de disolvente explícito y entornos de proteínas solvatadas usando simulaciones de GCMC iterativas. El método de la presente invención es un método computacional tal como se define en la reivindicación 1.

En determinadas realizaciones, los métodos de la presente invención pueden generar además una salida de las distribuciones espaciales del uno o más solutos y agua. Entre otras cosas, el resultado puede usarse para ayudar con el diseño de fármacos asistido por ordenador.

Los detalles de la distribución espacial dependerán del número de repeticiones de las etapas 2) y 3), definiendo el usuario cuándo se obtiene una distribución espacial satisfactoria. Normalmente, un usuario deja de repetir las etapas 2) y 3) cuando los detalles de la distribución espacial del uno o más solutos y agua ya no experimentan un cambio significativo con cada repetición adicional según lo definido por el usuario. Esta situación puede denominarse una distribución espacial convergente. En una disolución acuosa, por ejemplo, una distribución espacial convergente será homogénea (es decir, la misma en todas partes para el agua y todos los solutos). Alternativamente, cuando el sistema es heterogéneo (es decir, contiene una macromolécula tal como se define en las reivindicaciones adjuntas) la distribución espacial de solutos y agua también será heterogénea. Sin embargo, la distribución espacial puede definirse como convergente si no experimenta cambios significativos con más repeticiones, donde el usuario define lo que es significativo.

En el contexto de la presente invención, el  $\mu_{ex}$  oscilante permite la determinación de la distribución espacial de los solutos y el agua y el término "convergencia" se refiere al número de etapas de GCMC en las que la distribución espacial del agua y los solutos ya no cambia. Mientras que  $\mu_{ex}$  todavía puede oscilar para cada muestreo de la distribución espacial, el valor de  $\mu_{ex}$  promedio, sin embargo, no puede cambiar una vez que se obtiene la convergencia. En una realización, la convergencia de distribuciones espaciales se identifica usando coeficientes de solapamiento. En una realización, las distribuciones espaciales se consideran convergentes cuando el coeficiente de solapamiento es mayor de 0,4. En una realización alternativa, la convergencia se identifica cuando la desviación de la concentración de un soluto en el sistema sometido a muestreo de la distribución espacial es sistemáticamente menor del 10% con respecto a la concentración objetivo durante al menos los últimos 20 ciclos de muestreo.

El sistema puede incluir solo agua, uno o más solutos o una combinación de agua y el uno o más solutos. La concentración objetivo puede ajustarse a cualquier nivel deseado. Sin embargo, en una realización particular que tiene solo agua y un soluto, la concentración objetivo del soluto se ajusta a 1 M y la concentración objetivo del agua se ajusta a 55 M.

El número de operaciones de GCMC en la etapa 2) puede ser mayor que uno para cada ciclo. En determinadas realizaciones la operación de GCMC se realiza desde aproximadamente 100 hasta 100.000 veces. Sin embargo, si es necesario pueden realizarse menos de 100 o más de 100.000 operaciones de GCMC. Las operaciones de GCMC pueden dividirse de cualquier modo entre el uno o más solutos y agua. En determinadas realizaciones, las operaciones de GCMC pueden dividirse por igual entre el uno o más solutos y agua. La proporción de operaciones de GCMC para cada uno del uno o más solutos también puede asignarse en cualquier razón deseada. Sin embargo, en determinadas realizaciones, la proporción de operaciones de GCMC se asigna basándose en la concentración objetivo de cada uno de los solutos y agua. En determinadas realizaciones, el  $\mu_{ex}$  del uno o más solutos y agua se aumenta y se disminuye alternativamente durante las operaciones de GCMC a través de ciclos consecutivos que implican las etapas 2) y 3). El proceso alternativo de aumentar o disminuir el  $\mu_{ex}$  del uno o más solutos y agua se repite hasta que la distribución espacial general ya no cambia ni alcanza un nivel de cambio aceptable.

Las macromoléculas también están presentes en los sistemas según la presente invención. Se seleccionan de una proteína, ARN, ADN, hidrato de carbono, lípido, compuesto químico orgánico, compuesto químico inorgánico, o cualquier combinación de los mismos. Dichas macromoléculas generalmente son mayores de 1000 Dalton, y menores de 10.000.000 Dalton. Generalmente, las macromoléculas están en un intervalo de 5.000 a 500.000 Dalton. En determinados sistemas que comprenden una macromolécula, el método de muestreo de la presente invención incluye además una etapa en la que el  $\mu_{ex}$  del uno o más solutos y agua se aumenta y se disminuye alternativamente durante las operaciones de GCMC a través de ciclos consecutivos que implican las etapas 2) a 3) una vez que la concentración

del uno o más solutos y agua alcanza sus valores objetivo. En realizaciones que comprenden una o más macromoléculas, la distribución espacial del uno o más solutos puede usarse para identificar la afinidad preferente de cada uno de los solutos por la macromolécula. Esto puede lograrse determinando la probabilidad relativa de que el uno o más solutos se ubiquen espacialmente adyacentes o en una de las macromoléculas frente a la otra o al resto del sistema. Lo mismo vale para múltiples sitios en una sola macromolécula. La afinidad de unión puede estimarse basándose en la probabilidad de que los solutos se encuentren en un sitio en relación con el resto del sistema o en relación con otro sitio en la propia macromolécula o frente a la otra macromolécula.

Aún en otras realizaciones, se muestrea la distribución espacial que sigue a la etapa 2) usando simulaciones de dinámica molecular (MD). Una simulación de MD después del GCMC permite tanto el muestreo conformacional de los solutos como el muestreo configuracional del sistema acuoso y la macromolécula. Para lograr una convergencia satisfactoria según lo definido por el usuario, se repite el proceso de GCMC-MD a través de múltiples iteraciones, oscilando de manera sistémica el  $\mu_{ex}$  de las especies estudiadas sobre las iteraciones para impulsar el intercambio de solutos y agua. Muchos de los ejemplos proporcionados en el presente documento se describen usando GCMC-MD, pero la simulación de MD puede no usarse en todas las realizaciones de la presente invención.

En determinadas realizaciones, el sistema usado en los métodos descritos en el presente documento es un sistema esférico finito. La diferencia entre los radios de los sistemas puede ser de 0,1 Å a 1000 Å. En determinadas realizaciones, la diferencia entre los radios es o es aproximadamente de 5 Å. Aunque los ejemplos específicos proporcionados en el presente documento se enseñan con un sistema esférico, puede usarse cualquier región definida del espacio.

A continuación se describe la teoría del GCMC. Hay cuatro posibles movimientos de GCMC en una molécula, M (es decir, soluto o agua): Inserción: lleva M al sistema A desde el depósito; deleción: retira M del sistema A y lo mueve de nuevo al depósito; traslación y rotación: M se traslada/rota dentro de un subvolumen que rodea la ubicación original de M en el sistema A. Las probabilidades de estos movimientos que se rijan según los criterios de Metropolis son:

$$\begin{aligned}
 P_{\text{insertar}} &= \min \left\{ 1, \frac{f_n}{n+1} e^{-\beta \Delta E} \right\} \\
 P_{\text{eliminar}} &= \min \left\{ 1, \frac{n}{f_{n-1}} e^{-\beta \Delta E} \right\} \\
 P_{\text{tras/rot}} &= \min \left\{ 1, e^{-\beta \Delta E} \right\} \\
 \text{donde } B &= \beta \mu_{ex} + \ln \bar{n} \quad \text{y} \quad \bar{n} = \bar{\rho} \bar{v} \quad (1)
 \end{aligned}$$

$\mu_{ex}$  es el exceso de potencial químico,  $\bar{n}$  es el número esperado de moléculas,  $\bar{\rho}$  es la densidad,  $\bar{v}$  es el volumen del sistema A,  $f_n$  es el volumen fraccional del subespacio donde se realizan los intentos de inserción,  $\Delta E$  es el cambio de energía debido a un movimiento,  $\beta$  es  $1/k_B T$ ,  $k_B$  es la constante de Boltzmann y  $T$  es la temperatura (300 K en el presente estudio). A través de la simulación de GCMC, el volumen del sistema de simulación A, la energía total y el número total de partículas entre el sistema A y su depósito normalmente son fijos. Sin embargo, en determinadas realizaciones, la energía total y el número total de partículas entre el sistema A y su depósito pueden variar.

Como se observa en la ecuación 1, el número esperado de moléculas de cada soluto " $\bar{n}$ " se calcula a partir de la concentración objetivo de cada soluto ( $N_{obj}$ ), y el volumen del sistema A. Por tanto, las interacciones de cada soluto con las otras moléculas en el sistema de simulación A ( $\Delta E$ ) y el  $\mu_{ex}$  suministrado determina las poblaciones de soluto de las simulaciones de GCMC. Puesto que las probabilidades de movimiento de las aguas o los solutos individuales se impulsan por sus valores de  $\bar{n}$  y  $\mu_{ex}$ , cuando el  $\mu_{ex}$  de un soluto es menor que el trabajo necesario para mover una molécula desde el depósito de fase gaseosa hasta el sistema A, se producirá una disminución en la concentración del soluto del sistema A. Del mismo modo, cuando el  $\mu_{ex}$  suministrado es más que el trabajo necesario, se producirá un aumento en la concentración de soluto en el sistema A. En consecuencia, el valor de  $\mu_{ex}$  suministrado a la simulación de GCMC puede variar según el número de partículas en el sistema de simulación y el número esperado de moléculas de cada soluto  $\bar{n}$ .

Con  $\bar{n}$  usado como objetivo para cada soluto para el cálculo de GCMC, se define que el sistema de simulación A contiene agua en concentraciones de fase en volumen de 55 M. A través de las simulaciones de GCMC-MD iterativas, se usa el número de solutos y el agua en el sistema A frente a su número ( $\bar{n}$ ) esperado como guía para variar sus respectivos  $\mu_{ex}$  a través de cada iteración posterior de GCMC. Las simulaciones comienzan con el valor de  $\mu_{ex}$  para solutos y agua ajustados a 0 en la primera iteración. En cada iteración posterior o subconjunto de iteraciones, se varía el valor de  $\mu_{ex}$  para solutos y agua en una magnitud que se rige por la desviación de los solutos y el agua en el sistema A con respecto a su número objetivo respectivo  $\bar{n}$ . A medida que las concentraciones de los solutos y las aguas alcanzan su número objetivo, disminuye la anchura de la variación de cada  $\mu_{ex}$  y esto define el inicio de la convergencia. Por tanto, en las simulaciones de estado convencional con un solo tipo de soluto a una concentración

de 1 M en agua, puede calcularse la energía libre de hidratación (HFE) a partir del valor de  $\mu_{ex}$  en convergencia. La oscilación de los valores de  $\mu_{ex}$  para agua y solutos continúa después de la convergencia para facilitar el muestreo espacial de solutos en condiciones de equilibrio.

- 5 El esquema 1 de la figura 1 ilustra una simulación de estado convencional en la que el sistema A comprende un solo tipo de soluto a una concentración de 1 M. El esquema II ilustrado en la figura 1, sin embargo, representa una simulación en la que el sistema comprende una mezcla acuosa de múltiples solutos. Tal como se ilustra en la figura 1, el sistema A, es una región esférica de radio,  $r_A$ , en el que se realizan movimientos de GCMC. Tal como se ilustra adicionalmente en esta figura, depósitos separados de solutos y agua están acoplados al sistema A. El sistema A también está sumergido en un sistema B más grande, que incluye agua adicional. Para los sistemas acuosos en volumen, el sistema B es una esfera más grande de radio  $r_B = r_A + dr$ . En una realización,  $dr = 5 \text{ \AA}$ . El sistema B más grande limita los efectos de borde, tal como los que surgen de los solutos hidrófobos que ocupan el borde del sistema A (figura 6). Alternativamente, el sistema B puede tratarse usando condiciones de contorno periódicas (PBC) y/o incluir otras entidades químicas además del agua. La figura 10 ilustra un sistema T4-L99A a modo de ejemplo para simulaciones de GCMC-MD. En esta figura, un sistema periódico B está sumergido en un sistema esférico A.

El proceso de GCMC-MD iterativo se realiza de la siguiente manera:

- 1) Realizar  $i$  etapas de GCMC para intercambiar solutos y aguas entre sus respectivos depósitos y el sistema de simulación A. Las  $i$  etapas se dividieron entre cada uno de los  $M$  solutos y agua. En el presente estudio se intentaron

movimientos de  $\frac{i}{2}$  y  $\frac{i}{M+1}$  GCMC (inserción/delección/traslación/rotación) para los solutos ( $F_1, F_1, F_2, \dots, F_M$ ) y el agua, en el esquema I y el esquema II, respectivamente. Sin embargo, no es necesario que el número de movimientos de GCMC para los diferentes solutos y agua sea igual. Se aleatoriza el orden en el que se intentan los cuatro movimientos GCMC posibles y la molécula (solutos o agua) en la que se realiza el movimiento. En la primera iteración, el valor de  $\mu_{ex}$  para solutos y agua se ajustó a 0. Los radios de la(s) esfera(s) de agua y soluto sometida(s) a movimientos de GCMC se ajustaron a  $r_A$  a través del proceso de GCMC, aunque las interacciones energéticas asociadas con los movimientos también incluyen contribuciones de cualquier molécula de agua u otras entidades químicas, fuera de la esfera de GCMC que se define como el sistema B.

- 2) Después de GCMC, se realizaron  $j$  etapas de MD en los sistemas combinados A y B. Para los sistemas esféricos finitos, los solutos se mantuvieron dentro de las dimensiones esféricas del sistema A, ( $r_A$ ) utilizando varios métodos conocidos. En una realización a modo de ejemplo, esto se realizó mediante el uso de restricciones esféricas armónicas de fondo plano tal como se describe por Coleman C, *et al.*, (2011), Proc Natl Acad Sci USA 108 (17): 6838-6842. Además, el agua y cualquier otra molécula en el sistema B que engloba el sistema A se sometieron a i) restricciones esféricas armónicas de fondo plano con  $r_B = r_A + dr$  cuando el sistema B es esférico o ii) condiciones de contorno periódicas.

- 3) En la etapa 3, el valor de  $\mu_{ex}$  para soluto y agua, ( $\mu_P$  ( $P = F_1, F_2, \dots, F_M$  solutos o W, agua) se varió en  $d\mu_P$ . Este nuevo valor de  $\mu_P$  se usó en la siguiente iteración de GCMC. La magnitud de  $d\mu_P$  se determinó por la desviación del número actual de solutos/agua,  $N_P$  ( $P = F_1, F_2, \dots, F_M$ , o W) en el sistema A a partir del número esperado derivado de la concentración objetivo de solutos/agua ( $N_P^{\text{objetivo}}$ ) derivado del número esperado de solutos/agua.

$$\mu_P = \mu_P + d\mu_P$$

donde

$$d\mu_P = d\mu_P \times 5, \text{ cuando } N_P = 0; \text{ para el ciclo } 1, |d\mu_P| = 0,5$$

$$d\mu_P = \begin{cases} d\mu_P \times \frac{N_P^{\text{objetivo}}}{N_P}, & N_P < 0,7 N_P^{\text{objetivo}} \\ \text{aleat}((d\mu_P - 0,5), (d\mu_P + 0,5)), & 0,7 N_P^{\text{objetivo}} < N_P < 2 N_P^{\text{objetivo}} \\ -d\mu_P \times \frac{N_P}{N_P^{\text{objetivo}}}, & N_P > 2 N_P^{\text{objetivo}} \end{cases} \quad (2)$$

- 45 Con la variación de  $d\mu_P$  dependiendo del sistema, pueden generarse diversos esquemas (estado convencional, mezcla acuosa, sistemas heterogéneos) usando la ecuación 2. Cabe señalar que pueden aplicarse diferentes esquemas para variar  $d\mu_P$ .

4) Se usó entonces nuevo valor de  $\mu_{\text{ex}}$  procedente de la etapa 3 de la iteración actual para realizar  $k$  iteraciones (etapas 1-3), de una simulación de GCMC-MD.

5 Las simulaciones de GCMC se realizaron usando un código interno de C++ que implementó el esquema GCMC basado en cuadrícula con el algoritmo de sesgo de cavidad para impulsar los intercambios de soluto y agua entre sus depósitos y el sistema acuoso A. Se obtuvieron los parámetros empíricos de campo de fuerza del soluto a partir de CGenFF, el campo de fuerza general en CHARMM y TIP3P, un modelo para agua usado en simulaciones de GCMC y MD. Las moléculas de soluto elegidas para el presente estudio representan diferentes funcionalidades químicas, incluyendo moléculas apolares tales como benceno y propano, moléculas polares neutras tales como acetaldehído, metanol y formamida, y moléculas cargadas de manera negativa y positiva tales como acetato y metilamonio, respectivamente.

15 Para impedir la agregación de solutos hidrófobos y cargados, y promover una convergencia más rápida, se introdujo un término de energía repulsiva solo entre los pares moleculares de benceno:benceno, benceno:propano, propano:propano, acetato:acetato, acetato:metilamonio y metilamonio:metilamonio. Esto se logró añadiendo una partícula sin masa al centro de masas del benceno y al carbono central del propano, el acetato y el metilamonio. Cada una de esas partículas no interactúa con ningún otro átomo en el sistema, sino sólo con otras partículas en las moléculas hidrófobas o cargadas a través del término de campo de fuerza de Lennard-Jones (LJ) usando los siguientes parámetros:  $\epsilon = 0,01$  kcal/mol; y  $R_{\text{min}} = 12,0$  Å. Todos los términos de campo de fuerza de LJ y las interacciones culombimétricas se calcularon durante GCMC (es decir, no hubo truncamiento de interacciones no unidas), incluyendo las interacciones con el sistema B. Las simulaciones se inician con un sistema vacío A. El proceso de GCMC aleatorizado con los solutos y las aguas se realizaron inicialmente en múltiplos de 50.000 movimientos hasta que las moléculas de agua en el sistema alcanzan una concentración en volumen de 55 M. En cada punto de este proceso iterativo de 50.000 movimientos de GCMC, el  $\mu_{\text{ex}}$  de los solutos y el agua se aumentó en 1 kcal/mol para acelerar la acumulación de agua y soluto en el sistema A. Durante este proceso, la concentración de los solutos puede aumentar más allá de sus valores objetivo. Después de alcanzar la concentración de agua en volumen de 55 M, se realizó el procedimiento GCMC-MD tal como se describió anteriormente en las etapas 1-4 después de reajustar el  $\mu_{\text{ex}}$  de los fragmentos respectivos y el agua a 0. Más de 50 iteraciones del GCMC-MD en las que se variaron los valores de  $\mu_{\text{ex}}$ , las concentraciones de solutos se acercan a sus respectivos valores objetivo. Para cada iteración, se realizaron 50.000 y 100.000 movimientos de GCMC para el esquema I y el esquema II, respectivamente.

35 Se usó el paquete GROMACS (versión 5.0) para todas las simulaciones de MD. Para los sistemas acuosos, para cada iteración, se simulaban los sistemas combinados A y B, incluyendo todos los solutos y las aguas, durante 500 ps usando un integrador de salto de rana (integrador GROMACS "md"), que tiene un lapso de tiempo de 2 fs, a 300 K a través de un termostato Nose-Hoover. Se usó el algoritmo LINCS para restringir las geometrías del agua y todos los enlaces covalentes que implican un átomo de hidrógeno, mientras que las interacciones de van der Waals (vdW) y electrostáticas se desconectaron suavemente en un intervalo de 8-10 Å. Los solutos y las moléculas de agua se mantuvieron dentro de las dimensiones esféricas de los sistemas A o B mediante la aplicación de restricciones armónicas de fondo plano con una constante de fuerza de  $1,2$  kcal/mol/Å<sup>2</sup> en los siguientes átomos de soluto o agua: la partícula sin masa en el centro geométrico de benceno, propano, acetato y metilamonio, los átomos de carbono de acetaldehído, metanol y formamida y el átomo de oxígeno del agua. Para los sistemas acuosos en volumen, el radio,  $r_B$  de las restricciones armónicas de fondo plano usado para definir el sistema B se aplicó solo al agua y fue 5 Å mayor que el radio  $r_A = 20$  Å definiendo la restricción aplicada a los solutos en el sistema A.

45 Para el sistema T4-L99A, se usaron las coordenadas 181L de PDB con el ligando de benceno después de la delección del ligando. Brevemente, se insertó la estructura de T4-L99A en cajas que replican los sistemas del esquema I y del esquema II que contienen 1 M de benceno y 0,25 M de cada uno de los diferentes fragmentos (benceno, propano, acetaldehído, metanol, formamida, acetato y metilamonio), respectivamente. De manera similar a los sistemas acuosos, se construyeron 10 de tales cajas para el esquema I con benceno 1 M y el esquema II con los múltiples solutos, insertándose los solutos aleatoriamente en cada una de las cajas. Estos sistemas se minimizaron en más de 1000 etapas iterativas usando el algoritmo de descenso más pronunciado en presencia de condiciones de contorno periódicas (PBC). Entonces se equilibraron los sistemas equilibraron durante 250 ps mediante la reasignación periódica de velocidades. Se usó la versión de salto de rana del integrador Verlet con un lapso de tiempo de 2 fs para calentamiento y equilibrio. Las interacciones electrostáticas de largo alcance se manejaron utilizando el método Ewald de malla de partículas con un punto de corte de espacio real de 12 Å. Se aplicó una función de conmutación a las interacciones de Lennard-Jones a 12 Å, y se aplicó una corrección isotrópica de largo alcance al componente de presión para las interacciones de Lennard-Jones más allá de la longitud de punto corte de 12 Å. Durante la minimización y el equilibrio, se aplicaron restricciones armónicas de posición con una constante de fuerza de  $2,4$  kcal/mol/Å<sup>2</sup> a proteínas sin átomos de hidrógeno. Para las simulaciones de MD en el protocolo iterativo, se eliminaron las restricciones de posición y se reemplazaron por restricciones débiles aplicadas solo a los átomos de carbono C-alfa de la estructura principal de la proteína con una constante de fuerza ( $k$  en  $1/2 k \Delta x^2$ ) de  $0,12$  kcal/mol/Å<sup>2</sup>. Esto se hizo para impedir la rotación de la proteína en la caja de simulación y para impedir la desnaturalización potencial de la proteína debido a la presencia de una disolución de fragmentos altamente concentrada (17).

65 Se repiten las iteraciones de GCMC-MD hasta que se obtenga la convergencia según lo definido por el usuario.

Normalmente, se necesitan aproximadamente 200 iteraciones de GCMC-MD para los sistemas acuosos, produciendo 100 ns de simulaciones de MD y 10 y 20 millones de etapas de MC para el esquema I y el esquema II, respectivamente. Además, para garantizar un muestreo y una convergencia suficientes, se realizaron 10 simulaciones de GCMC-MD independientes para el esquema II y para cada soluto en el esquema I, dando como resultado 1  $\mu$ s (10  $\times$  100 ns) acumulativo de MD para el esquema II y cada soluto en el esquema I. Las ejecuciones difieren a través de un origen generado aleatoriamente tanto para GCMC como para las operaciones de integrador de MD de salto de rana en cada iteración. Los GCMC-MD de los sistemas de proteínas se repiten en más de 100 iteraciones, produciendo 500 ns acumulativos de MD a lo largo de las 10 simulaciones independientes de los sistemas tanto del esquema I como del esquema II. En el contexto del esquema I, las simulaciones de GCMC-MD permitirían el cálculo de la HFE correspondiente a un sistema acuoso de estado convencional de 1 M y la afinidad de unión de solutos a la proteína. Con la mezcla de solutos en el esquema II, el enfoque permitiría determinar el  $\mu_{ex}$  requerido para impulsar el muestreo de la distribución de moléculas de soluto en un sistema acuoso heterogéneo y determinar los patrones de afinidad de soluto alrededor del sitio de proteína.

En los cálculos iniciales, cuando el soluto y el disolvente en los sistemas esféricos finitos compartían el mismo límite esférico, las moléculas no polares solo muestreaban la superficie del sistema esférico (figura 6). Esto se debe a las interacciones favorables de los solutos no polares con el entorno de vacío no polar fuera del sistema de simulación esférico. Para superar esto, el radio del sistema de simulación para el agua se aumentó en 5 Å, de modo que se evita el “aplanamiento” de las moléculas no polares ya que permanecen completamente hidratadas. El uso de un sistema extendido de este tipo puede aliviar otros posibles efectos de borde que pueden encontrar otros solutos. En un aspecto de la invención, el sistema B más grande puede extenderse a cualquier distancia más allá de las restricciones de solutos que definen el sistema A. El sistema B más grande también puede modelarse usando las condiciones de contorno periódicas (PBC), que se describen en los cálculos para el sistema T4-L99A.

Se consideraron dos tipos de sistemas acuosos: 1) un sistema que contiene solo un tipo de soluto en agua a una concentración de 1 M, replicando de ese modo el estado convencional del soluto y 2) una mezcla acuosa diluida que contiene muchos tipos de solutos cada uno a una concentración de 0,25 M. Los resultados que usan ambos esquemas I y II se presentan para sistemas acuosos solo, seguidos de una descripción para los cálculos del esquema I y del esquema II en un sistema que contiene la lisozima T4-L99A. Este mutante de lisozima se seleccionó como sistema modelo, ya que se ha usado ampliamente en estudios experimentales y computacionales de unión de ligandos, así como en estudios que evalúan el impacto de las mutaciones en la estructura y estabilidad de la proteína. La mutación L99A en el dominio C-terminal de T4 crea una cavidad hidrófoba completamente enterrada de  $\sim 150 \text{ \AA}^3$  que, aunque inaccesible en estructuras estáticas, une pequeños ligandos hidrófobos de manera rápida y reversible. Una cavidad ocluida de este tipo ofrece una prueba rigurosa de la eficacia de muestreo de la metodología GCMC-MD presentada, incluyendo una evaluación cuantitativa del enfoque.

Brevemente, la metodología GCMC-MD puede explicarse de la siguiente manera. Se realizó GCMC en agua y en cada uno de los diferentes solutos en el sistema. Se usaron variaciones en el valor de  $\mu_{ex}$  para agua y solutos para las iteraciones de GCMC-MD para mejorar las probabilidades de intercambio de solutos y permitir simultáneamente la determinación del  $\mu_{ex}$  requerido para mantener una concentración definida de los solutos y las moléculas de agua en los sistemas acuosos. Este enfoque, por tanto, permite determinar la HFE de los solutos usando sus valores de  $\mu_{ex}$  cuando la concentración objetivo de soluto en los sistemas acuosos es de 1 M. En presencia de la proteína, se usó la estrategia de GCMC-MD para determinar las afinidades de unión de los solutos a la proteína, así como para muestrear eficazmente la distribución espacial de los solutos en y alrededor de la proteína. La identificación de sitios mediante la metodología de saturación competitiva de ligandos (SILCS) es un método de muestreo basado en fragmentos que mapea patrones de afinidad de energía libre de grupos funcionales en las superficies de proteínas, incluyendo las cavidades ocluidas de unión a ligando de las proteínas. El método de SILCS, por tanto, permite la determinación de los patrones de afinidad de múltiples solutos a la proteína y puede usarse para el diseño racional de fármacos.

## Métodos

I. Detalles de la simulación para el esquema I y el esquema II de sistemas acuosos: dinámica molecular con condiciones de contorno periódicas

Los parámetros de campo de fuerza empíricos para las proteínas son CHARMM36, las moléculas de ligando se trataron usando CGenFF y el agua se trató usando el modelo TIP3P. Las simulaciones se realizaron usando el programa de simulación molecular CHARMM. Para replicar el esquema I, el número de moléculas requeridas para alcanzar una concentración de 1M de un soluto específico se colocaron aleatoriamente en una caja cúbica cuyos lados tenían 50 Å cada uno. De manera similar, para replicar el esquema II, el número de moléculas requeridas para alcanzar una concentración de 0,25 M de cada tipo de soluto, se colocaron aleatoriamente en una caja cúbica de agua en volumen cuyos lados tenían 50 Å cada uno. Las mezclas acuosas se minimizaron usando el algoritmo de descenso más pronunciado en 5000 etapas y se minimizaron adicionalmente en otras 5000 etapas usando el algoritmo de gradiente conjugado en presencia de condiciones de contorno periódicas. Entonces se calentaron los sistemas hasta 300 K a una velocidad de 5 K/ps mediante reasignación periódica de velocidades. Después de esto, el sistema se equilibró durante 200 ps usando la reasignación de velocidad. Se usó la versión de salto de rana del integrador Verlet

con un lapso de tiempo de 1 fs para calentamiento y equilibrio. Las geometrías de agua y los enlaces covalentes que implicaban átomos de hidrógeno se restringieron utilizando el algoritmo SHAKE. Las interacciones electrostáticas de largo alcance se manejaron con el método Ewald de malla de partículas con punto de un corte de espacio real de 12 Å. Se aplicó una función de conmutación a las interacciones de Lennard-Jones a 10 Å, y se aplicó una corrección isotrópica de largo alcance al parámetro de presión para las interacciones de Lennard-Jones más allá de la longitud de corte de 12 Å. Después del equilibrio, las cajas de mezcla acuosa se simularon durante 15 ns con un lapso de tiempo de 2 fs, a 300 K y a 1 atm de presión usando un termostato Nose-Hoover y el barostato de pistón de Langevin.

## II. Generación de FragMaps

Las distribuciones de probabilidad 3D para átomos seleccionados de solutos a partir de las simulaciones de GCMC/MD y GCMC solo se denominan "FragMaps". Para la presente invención, se construyeron FragMaps para carbonos de benceno, carbonos de propano, hidrógeno polar de metanol, oxígeno de metanol, hidrógenos polares de formamida, oxígeno de formamida, nitrógeno no protonado de imidazol, hidrógeno de imidazol, oxígeno de acetaldehído, hidrógenos polares de metilamonio y oxígenos de acetato. Brevemente, los átomos de soluto de instantáneas de 10 ps de los sistemas de mezcla acuosa del esquema II en ausencia y en presencia de la simulación de GCMC-MD del mutante T4-L99A se agruparon en elementos de volumen cúbico (vóxeles)  $1 \text{ \AA} \times 1 \text{ \AA} \times 1 \text{ \AA}$  de una cuadrícula que abarca todo el sistema y se calculó la ocupación de vóxel para cada tipo de átomo de FragMap.

Para las simulaciones sólo de GCMC, se obtuvieron ocupaciones de vóxel para cada tipo de átomo de FragMap cada 1000 etapas de GCMC en cada ciclo. Estos se normalizaron entonces por las ocupaciones de vóxel de los fragmentos en un sistema de fase en volumen desprovisto de la proteína. Las ocupaciones de fase en volumen fueron iguales en las simulaciones de GCMC/MD y solo de GCMC. Las ocupaciones de vóxel de los once tipos de átomos se fusionaron para crear los siguientes cinco tipos genéricos de FragMap: (1) compuestos no polares genéricos, APOLAR (carbonos de benceno y propano); (2) donador de enlace de hidrógeno neutro genérico, HBDon (hidrógenos polares de metanol, formamida e imidazol); (3) aceptor de enlace de hidrógeno neutro genérico, HBACC (oxígeno de metanol, formamida y acetaldehído y nitrógeno no protonado de imidazol) (4) donador positivo, POS (hidrógenos polares de metilamonio); y (5) aceptor negativo, NEG (oxígenos de acetato).

Se normalizaron las distribuciones de probabilidad para cada tipo de átomo en ausencia de proteína y luego se convirtieron en energías libres a través de una transformada basada en Boltzmann de la probabilidad normalizada de producir una energía libre en la cuadrícula (GFE) para cada tipo de fragmento "f" para las coordenadas x, y, z, denominados FragMaps de GFE. Todos los valores de GFE se limitaron a 3 kcal/mol. Los FragMaps se visualizaron como superficies de iso-contorno a un valor de GFE de -1,2 kcal/mol, a menos que se indique de otro modo.

## III. Puntuación de energía libre en la cuadrícula de ligandos (LGFE)

La LGFE cuantifica el solapamiento de átomos en ligandos en la cavidad de unión de ligando (LBP) con los correspondientes FragMaps de GFE. Los átomos de ligando se clasificaron en tipos de FragMap, según un mapa de asignación. Brevemente, a cada átomo clasificado de un ligando con coordenadas  $(x_i, y_i, z_i)$  se le asignó una puntuación igual al valor de GFE del tipo de FragMap correspondiente (f),  $GFE_{x_i, y_i, z_i}^f$ , del vóxel que ocupa. La LGFE es entonces la suma de cada uno de estos valores de GFE para todos los átomos de ligando clasificados. La LGFE se calculó como el promedio ponderado de Boltzmann sobre un conjunto de conformaciones obtenidas por muestreo de MC del ligando en la LBP, en el campo de FragMaps.

## IV. Conjunto conformacional de MD (para $LGFE^{MD}$ ):

Las conformaciones iniciales del ligando se extrajeron a partir de los correspondientes archivos PDB de coordenadas co-cristalinas y se utilizó el algoritmo de parametrización CGenFF automatizado para obtener la topología y los parámetros en el contexto de CGenFF. Se extrajo la conformación de ligando a partir de la estructura co-cristalina y se alineó con la conformación de proteína usada en las simulaciones de SILCS. La alineación se realizó basándose en la alineación óptima de los átomos  $C_\alpha$  de la estructura principal en las dos estructuras de proteínas. El complejo se minimizó usando el algoritmo SD para 50 etapas. Esta conformación minimizada se sometió a una simulación de MD de 10 ps, con instantáneas emitidas cada 0,2 ps. Durante la dinámica, se restringieron todos los átomos de proteínas más allá de 8 Å alrededor de cualquier átomo de ligando usando una constante de fuerza de 1 kcal/mol/Å<sup>2</sup>. Este proceso se repitió en 40 conformaciones de proteínas obtenidas a partir de las simulaciones de GCMC-MD SILCS, espaciadas por igual en el tiempo (ciclo 10, 20, 30, ...), produciendo un total de 1000 conformaciones de ligandos para la estimación de la LGFE ponderada de Boltzmann.

## Conjunto conformacional de MC (para $LGFE^{MC}$ ):

El muestreo del ligando también se realizó en el "campo" de los Fragmaps de GFE usando las etapas de Metropolis - Monte Carlo (MC). Se usó un conjunto interno de programas para configurar y realizar las simulaciones de MC. El ligando tenía grados de libertad rotacionales, traslacionales e intramoleculares. El ligando no tenía restricciones rotacionales, pero su centro de masa (CoM) estaba restringido a encontrarse dentro de 2,5 Å del CoM de la

conformación cristalina del ligando usando una restricción de fondo plano. Los grados de libertad intramoleculares consistieron en enlaces rotativos, que se detectaron automáticamente basándose en la topología molecular de CGenFF. Todos los enlaces no terminales acíclicos se consideraron rotativos, con la excepción de los enlaces que terminan en grupos metilo o  $\text{NH}_3^+$ . Los términos de campo de fuerza correspondientes a los grados de libertad intramoleculares comprendían los términos diédrico, van der Waals (vdW) y electrostático. Debido a la ausencia de proteínas y disolventes durante estas simulaciones, se usó un dieléctrico dependiente de la distancia ( $= 4|r|$ ) para evaluar las contribuciones electrostáticas intramoleculares para evitar su sobreestimación. La energía calculada durante Metropolis – MC puede escribirse de la siguiente manera.

$$E = E_{\text{vdw, intra}} + E_{\text{elec, intra}} + E_{\text{dihe, intra}} + \text{LGFE}$$

Para cada ligando, se realizaron 20 simulaciones de MC diferentes (cada una de ellas para 10.000.000 etapas; instantáneas registradas cada 10.000 etapas), donde para cada una de ellas el ligando se coloca inicialmente de manera aleatoria cerca del sitio activo de la proteína. Primero se calcula una LGFE promedio sobre cada una de estas 20 simulaciones de MC. La  $\text{LGFE}^{\text{MC}}$  es entonces el promedio ponderado de Boltzmann sobre estos valores de LGFE obtenidos a través de las 20 simulaciones de MC.

#### V. Determinación de la capacidad de oclusión de la cavidad de unión de ligando

El grado en que los sitios de unión están ocluidos se denomina capacidad de oclusión, y se calcula como la razón del área superficial accesible al disolvente (SASA) de los ligandos solos con respecto al SASA cuando se une a la proteína usando estructuras de rayos X del complejo ligando-proteína. Tal como se muestra en la tabla 1, estas razones varían desde 0, para la testosterona unida a AR hasta 0,13, tanto para la rosiglitazona unida a  $\text{PPAR}_\gamma$  como para el carazalol unido a  $\beta_2\text{AR}$ . Estos datos sugieren un alto nivel de inaccesibilidad de las LBP para estas dianas de fármaco terapéuticamente importantes.

Tabla 1

Receptor	Ligando	SASA en volumen	SASA de LBP	Razón
AR	Testosterona	402,1	0	0
$\text{PPAR}_\gamma$	Rosiglitazona	506,8	68,6	0,13
	Ácido decanoico	341,1	28,8	0,08
mGluR	FITM	576,8	43,8	0,07
$\beta_2\text{AR}$	BI-167107	534,8	36,32	0,07
	Carazalol	478,6	62,8	0,13

#### VI. GCMC-MD para mapear patrones de energía libre de grupos funcionales en LBP ocluidas

Durante el GCMC, los fragmentos y el agua se intercambian entre sus depósitos de fase gaseosa y una región cúbica de radio 20 Å (25 Å para  $\text{PPAR}_\gamma$  y AR) que engloba la cavidad de unión de ligando de la proteína. El exceso de potencial químico ( $\mu_{\text{ex}}$ ) suministrado para impulsar los intercambios de fragmentos fluctúa periódicamente, por ejemplo, cada 3 ciclos, de manera que el  $\mu_{\text{ex}}$  promedio durante los 100 ciclos está cerca de los valores mostrados en la tabla 2. Estos valores son la magnitud de  $\mu_{\text{ex}}$  requerida mantener 0,25 M de un soluto en una mezcla acuosa en volumen desprovista de cualquier proteína, y son aproximadamente iguales a la energía libre de hidratación.

Tabla 2

Fragmento	HFE <sup>sep</sup> 31 (kcal/mol)	$\mu_{\text{ex}}$ (kcal/mol)
Benceno	-0,71	-0,79
Propano	1,60	1,96
Acetaldehído	-4,43	-3,23
Metanol	-6,16	-5,62
Formamida	-10,71	-10,92
Imidazol	-12,55	-14,18
Acetato	-96,5	-97,31
Metilamonio	-52,0	-68,49
Agua	-	-5,6

Se realizaron diez simulaciones de GCMC-MD para cada sistema de proteínas. Cada una de estas 10 simulaciones constituía 100 ciclos de GCMC y MD (200 en el caso de AR), implicando cada ciclo 100.000 etapas de GCMC y 0,5 ns de MD, produciendo un total de 100 millones de etapas de GCMC y 500 ns de MD para cada sistema de proteínas (200 millones de etapas de GCMC y 1  $\mu\text{s}$  de MD para AR). A través de los 100.000 etapas de GCMC, tanto las moléculas de soluto (fragmentos) como el agua se intercambian entre sus depósitos de fase gaseosa y un subvolumen que encapsula las LBP de la proteína. La configuración al final del GCMC se usa como configuración de partida en la

MD.

Antes de comenzar la MD, se realiza una minimización de descenso pronunciado (SD) de 500 etapas y un equilibrio de 100 ps. El exceso de potencial químico ( $\mu_{ex}$ ) suministrado para impulsar los intercambios de soluto y agua durante el GCMC se hacen fluctuar periódicamente durante cada 3 ciclos, de manera que el  $\mu_{ex}$  promedio a lo largo de los 100 ciclos está cerca de los valores de la tabla 2. Estos valores son la magnitud de  $\mu_{ex}$  requerida para mantener 0,25 M de un soluto en una mezcla acuosa en volumen desprovista de cualquier proteína, y son aproximadamente iguales a la energía libre de hidratación. Puesto que  $\mu_{ex}$  se hace fluctuar periódicamente, a lo largo de todo el GCMC, el sistema no es un conjunto formal de GC. Sin embargo, al mantener el  $\mu_{ex}$  promedio constante a lo largo de la duración de la simulación, el grado de desviación fue mínimo. Para todas las proteínas ejemplificadas en este método, se monitorizó la convergencia de los FragMaps calculando coeficientes de solapamiento (OC).

Las diez trayectorias se dividieron en dos grupos (grupo 1, trayectorias 1-5; grupo 2, trayectorias 6-10), y los FragMaps de cada grupo se calcularon por separado. El OC relaciona el solapamiento entre FragMaps de dos grupos (grupo 1, trayectorias 1-5 y grupo 2; trayectorias 6-10) con un número entre 0 y 1, reflejando 1 mapas completamente idénticos. El OC se calcula usando la ecuación (3).

$$OC = \sum_{i=1}^N \min \left( \frac{Q_i^1}{\sum_{j=1}^N Q_j^1}; \frac{Q_i^2}{\sum_{j=1}^N Q_j^2} \right) \quad (3)$$

En la ecuación 3, N es el número de vóxeles en los FragMaps y  $Q_i^1$  y  $Q_i^2$  son ocupaciones para el vóxel  $i$ ésimo a partir de los FragMaps generados de los grupos 1 y 2, respectivamente. Las razones entre paréntesis se calculan para normalizar la ocupación de cada vóxel dividido entre la suma de las ocupaciones de todos los vóxeles en el FragMap correspondiente. Para cada vóxel, los valores más pequeños (la parte conservada) de los grupos 1 y 2 se suman en todos los vóxeles para obtener el OC. Cabe señalar que el índice de OC no se comporta linealmente, de manera que una diferencia relativamente pequeña en las dos distribuciones conduce a una disminución desde 1 hasta aproximadamente 0,8, y los valores  $> 0,5$  indican un alto grado de similitud (figura 11). Tal como se muestra en la tabla 3, se registran coeficientes de solapamiento razonables para FragMaps de todos los sistemas sometidos a prueba.

Tabla 3

Receptor	APOLAR	POS	NEG	HBDon	HBACC
AR	0,68	0,65	0,63	0,41	0,47
PPAR $\gamma$	0,78	0,66	0,75	0,58	0,63
mGluR	0,75	0,52	0,55	0,55	0,49
B2A	0,73	0,79	0,49	0,53	0,51
B2I	0,77	0,73	0,62	0,49	0,62

#### VII. simulaciones sólo de GCMC

Para examinar el papel de la flexibilidad de la proteína en la modulación de FragMaps en las LBP, se realizó un segundo conjunto de simulaciones sólo de GCMC para todos los sistemas de proteínas. Al igual que la simulación de GCMC/MD, se configuraron 10 ejecuciones independientes. A través de los 100 ciclos, no se realizó MD al final de 100.000 etapas de GCMC. Sin embargo, la última configuración al final de los 100.000 etapas de GCMC se usó como la configuración de entrada para el siguiente ciclo. Al igual que con el GCMC-MD, el  $\mu_{ex}$  se hace fluctuar periódicamente durante cada 3 ciclos alrededor de los valores que mostrados en la tabla 2, anteriormente.

#### VIII. Preparación de proteínas

##### AR y PPAR $\gamma$ :

Se usaron las coordenadas de PDB de 2AM9 (resolución 1,64 Å) y 3U9Q (resolución 1,52 Å) con los ligandos testosterona (TES) y ácido decanoico (DA), respectivamente, para los métodos basados en ordenador de la presente invención. Después de la delección de los ligandos, los residuos faltantes 262-275 en 3U9Q de PDB se construyeron usando el programa MODELLER. Se generaron y clasificaron un total de 100 modelos usando el método de energía de proteína optimizada discreta (DOPE) y se usó el modelo de clasificación más alta como estructura de partida. Se retuvieron las moléculas de agua cristalina, al igual que cualquier ion estructuralmente importante. Un conjunto de instrucciones de preparación interna utilizó las utilidades GROMACS para generar el sistema de simulación que implica proteínas, agua y moléculas pequeñas, con el tamaño del sistema para tener la extremidad de la proteína separada del borde en 12 Å en todos los lados. La carga neta del sistema hizo neutra reemplazando moléculas de agua

aleatorias por el número apropiado de iones de sodio o cloruro. Las proteínas se minimizaron durante 500 etapas con el algoritmo de descenso más pronunciado (SD) en presencia de condiciones de contorno periódicas (PBC). A esto le siguió un equilibrio de 100 ps durante el cual la temperatura se ajustó mediante un cambio de escala de velocidad. Durante la minimización y el equilibrio, se aplicaron restricciones armónicas de posición con una constante de fuerza de 2,4 kcal mol<sup>-1</sup> Å<sup>-2</sup> a proteínas sin átomos de hidrógeno. Las coordenadas finales al final del equilibrio se utilizan como las conformaciones de partida para la simulación de GCMC/MD.

#### IX. Examen virtual (VS)

Se usó una base de datos *in silico* de aproximadamente 1,8 millones de compuestos que contenía todos los tautómeros accesibles y estados de protonación de cada compuesto a partir de las bases de datos CHEMBRIDGE y MAYBRIDGE para el examen. 1) Distintas afinidades de FragMap de B21 (A1, A2, P1, P2, HBD1 y HBD2 en la figura 12) se convirtieron en características de farmacóforo (SILCS-Pharm) usando un método que es una extensión del trabajo recientemente publicado. PHARMER se usó para anotar cada uno de los ligandos a partir de la base de datos con las características de SILCS-PHARM y luego combinarlas con las características de los FragMaps basados en la RMSD de una alineación de las características coincidentes. Un punto de corte de RMSD < 1,6 Å produjo 11.119 ligandos. Estos ligandos se acoplaron a las estructuras tanto activa (PDB: 3P0G) como inactiva (PDB: 2RH1) de β<sub>2</sub>AR usando Autodock Vina, y las diferencias en las puntuaciones de las conformaciones mejor clasificadas se calcularon como:

$$\Delta E_{\text{acop}} = \left| E_{\text{acop}}^{\text{act}} \right| - \left| E_{\text{acop}}^{\text{inact}} \right|$$

donde  $E_{\text{acop}}^{\text{act}}$  y  $E_{\text{acop}}^{\text{inact}}$  son puntuaciones de las posiciones mejor clasificadas en las estructuras activa e inactiva, respectivamente. Un segundo criterio de selección de  $\Delta E_{\text{acoplamiento}} > 0$ , produjo 906 ligandos. Las LGFE de estos 906 ligandos se calcularon como promedios de Boltzmann a lo largo de más de 1000 etapas de muestreo de MC en campos de FragMaps tanto activos ( $LGFE^{\text{act}}$ ) como inactivos ( $LGFE^{\text{inact}}$ ). Las diferencias en estas puntuaciones de LGFE se calcularon como  $\Delta LGFE = |LGFE^{\text{act}}| - |LGFE^{\text{inact}}|$ , y se seleccionaron 109 ligandos con  $\Delta LGFE > 0$ . Se usó un análisis de agrupamiento basado en huella química para seleccionar 15 ligandos químicamente diversos con los que se realizaron estudios de evaluación funcional.

#### Fisiología de vías respiratorias intactas *ex vivo*:

Todos los estudios en ratones se aprobaron por el Comité para Cuidado y Uso de Animales de UMB. Se cortaron secciones de 5 mm de tráquea de ratones FVB/N y se estudiaron en un sistema de miógrafo isométrico (ADInstruments) tal como se describe en la bibliografía. Se aplicó una tensión pasiva de 0,5 g para cada anillo para un nivel inicial. Los anillos se contrajeron con acetilcolina 1 μM, seguido de la adición de 100 μM de isoproterenol (iso) y los ligandos seleccionados de los estudios de VS. Se midió el porcentaje de relajación en presencia de un ligando como un cambio en la tensión desde la contracción máxima estimulada por acetilcolina. Los anillos se lavaron para su reutilización. La relajación se calculó como un promedio de más de 9 series con isoproterenol y 4 series con cada uno de los ligandos seleccionados.

### Ejemplos

#### I. Sistemas de disolución acuosa

Para el esquema I y la mezcla acuosa del esquema II, la concentración objetivo se obtiene variando  $\mu_{\text{ex}}$  a través de las iteraciones de GCMC-MD. Para comenzar, el valor de  $\mu_{\text{ex}}$  tanto para el soluto como para el agua se ajusta inicialmente a 0. En ambos casos, el valor de  $\mu_{\text{ex}}$  promedio tanto para solutos como para el agua convergió cerca de su HFE. La tabla 4 enumera tanto la HFE calculada como la experimental ( $HFE^{\text{exp}}$ ) de los solutos. Se obtuvieron los  $\mu_{\text{ex}}$  promedio de soluto y agua para las simulaciones de GCMC-MD del esquema I y el esquema II con un sistema esférico de radio 25 Å a partir de los ciclos 150-200. Los promedios y las desviaciones estándar se basaron en las 10 simulaciones individuales de cada sistema.

Tabla 4

Fragmento	HFE <sup>exp</sup> (kcal/mol)	HFE <sup>tep</sup> (kcal/mol)	Esquema I		Esquema II	
			$\mu_{\text{ex}}$ (kcal/mol)	Conc (M)	$\mu_{\text{ex}}$ kcal/mol)	Conc (M)
Benceno	-0,83	-0,86	-0,71±0,37	1,40±0,11	-0,94±0,13	0,32±0,19
Propano	1,96	1,89	1,35±0,12	1,37±0,26	1,34±0,52	0,36±0,11
Acetaldehído	-3,5	-2,87	-3,1±0,54	1,01±0,11	-2,86±0,16	0,24±0,12
Metanol	-5,1	-4,83	-5,79±0,23	1,3±0,49	-4,92±0,14	0,22±0,08
Formamida	-14	-9,12	-14,2±2,10	1,11±0,14	-12,33±2,20	0,20±0,09

Acetato	-79,1	-97,43	-48,1±0,54	0,82±0,21	-52,1±0,49	0,24±0,03
Metilamonio	-71,3	-60,02	-58,1±0,34	0,71±0,23	-56,1±0,52	0,22±0,13
Agua	-5,6	-	-5,2±0,09	53,7±1,3	-4,9±0,14	55,1±0,32

Las HFE también se calcularon usando un método de perturbaciones de energía rápidas (FEP), tal como se describió anteriormente (véase Baker CM, *et al.*, (2010), J Chem Theory Comput 6 (4): 1181-1198), para tener en cuenta las posibles limitaciones en el campo de fuerza que producirían una HFE en desacuerdo con los datos experimentales.

5 La figura 2 traza la progresión tanto de la concentración como de  $\mu_{ex}$  a través de las iteraciones de GCMC-MD, a

medida que  $\mu_{ex}$  se variaba basándose en  $N_P^{objetivo}$  en el sistema A tanto para el esquema I como para el esquema II. Tanto la concentración como el  $\mu_{ex}$  en cada iteración de GCMC-MD en la figura 2 se presentan como el promedio de más de 10 simulaciones independientes. Tanto en el esquema I como en el esquema II, los solutos y el disolvente del

10 sistema A alcanzan sus valores objetivo  $N_P^{objetivo}$ , correspondientes a una concentración de 1 M y 0,25 M de soluto en los esquemas I y II, respectivamente, y una concentración de 55 M para el agua. Para la mayoría de los sistemas, se produjo convergencia dentro de 50 iteraciones. Sin embargo, para los fragmentos cargados, las tasas de aceptación para las inserciones de partículas fueron bajas, debido a las interacciones electrostáticas desfavorables. Debido a que la convergencia tardó más en estos casos, se necesitaron aproximadamente 200 iteraciones de la simulación para lograr la convergencia.

15 La tabla 4 enumera el valor de  $\mu_{ex}$  promedio y la concentración para cada soluto para 10 simulaciones, a partir de las 50 iteraciones finales. Dado que el valor de  $\mu_{ex}$  promedio y la concentración para los solutos a lo largo de las últimas 50 iteraciones son muy similares a los valores de  $\mu_{ex}$  y concentración medidos a través de las iteraciones 50-200 (tabla 5) de la simulación, estos resultados concuerdan con el inicio de la convergencia a partir de la iteración 50.

20

Tabla 5

Fragmento	Esquema I		Esquema II	
	$\mu_{ex}$ (kcal/mol)	Conc (M)	$\mu_{ex}$ (kcal/mol)	Conc (M)
Benceno	-0,62±0,47	1,21±0,21	-0,94±0,23	0,32±0,21
Propano	1,39±0,23	1,32±0,13	1,31±0,21	0,37±0,13
Acetaldehído	-3,1±0,71	1,1±0,81	-2,92±0,43	0,24±0,12
Metanol	-5,73±0,31	1,1±0,34	-4,92±0,21	0,25±0,11
Formamida	-14,2±1,22	1,03±0,43	-11,51±1,31	0,22±0,12
Acetato	-50,2±0,71	0,83±0,31	-52,1±0,62	0,24±0,04
Metilamonio	-58,1±0,42	0,71±0,32	-56,1±0,74	0,22±0,17
Agua	-5,2±0,12	54,4±1,1	-5,0±0,13	55,1±0,41

25 Para moléculas hidrófobas y polares, los valores de  $\mu_{ex}$  promedio se comparan bien con la HFE<sup>fep</sup>. La desviación más grande se produce con formamida con los valores que se encuentran entre los valores de HFE<sup>exp</sup> y HFE<sup>fep</sup>. La desviación de  $\mu_{ex}$  y HFE<sup>fep</sup> con respecto a la HFE<sup>exp</sup> para fragmentos cargados, acetato y metilamonio, se debe al hecho de que no está teniendo en cuenta el potencial de superficie de contacto de vacío-disolvente en los presentes cálculos. Para aniones/cationes monovalentes, se calculó que esta contribución era de aproximadamente +/-12,5 kcal/mol, respectivamente, usando el modelo de agua de TIP3P. Además, dado que los métodos de GCMC con solutos cargados están limitados por bajas tasas de aceptación para inserciones de partículas, las presentes estimaciones de  $\mu_{ex}$  para sistemas cargados están dentro de límites aceptables. En conjunto, estos resultados establecen que la metodología de GCMC-MD iterativa presentada es útil frente a estimar la HFE de solutos orgánicos gracias a su  $\mu_{ex}$  en sistemas acuosos en estado convencional. Además, el enfoque es adecuado para mezclas acuosas más complejas tal como se demuestra por el hecho de que los valores de  $\mu_{ex}$  obtenidos a partir del esquema II concuerdan de manera

35

Aunque el protocolo de simulación de GCMC-MD logró la concentración y  $\mu_{ex}$  correctos en el sistema, se investigó el método para determinar su capacidad para obtener el muestreo espacial correcto de solutos en sistemas esféricos finitos, por ejemplo, sistemas heterogéneos. Se investigó el muestreo espacial mediante el análisis de funciones de distribución radial (RDF). Se calcularon las RDF de átomos de soluto seleccionados y oxígenos de agua a partir del muestreo de MD acumulativo de simulaciones de GCMC-MD tanto del esquema I como del esquema II.

40

Se compararon estas RDF con las RDF obtenidas a partir de simulaciones de PBC-MD de 15 ns de agua explícita. Las simulaciones de PBC-MD tal como se describió anteriormente mantienen concentraciones de solutos y agua según los parámetros expuestos para el esquema I y el esquema II en una caja cúbica cuyos lados tienen 50 Å cada uno. Estas simulaciones también incluyen un tratamiento explícito para interacciones no de unión de largo alcance mediante métodos de corrección de LJ isotrópica y de Ewald de malla de partículas.

45

Las RDF a partir de muestreo de GCMC-MD de sistemas esféricos finitos y las simulaciones de PBC-MD de agua explícita coinciden muy bien. Estos resultados indican que el uso de valores de  $\mu_{ex}$  oscilantes para impulsar el muestreo de GCMC y que el tratamiento de las interacciones no de unión de largo alcance en el protocolo de GCMC-MD presentado no tienen un impacto significativo sobre el muestreo espacial de los sistemas acuosos. Se observan algunas fluctuaciones en la RDF a partir de las PBC de esquema II, que se deben a cuestiones de muestreo, por ejemplo, las simulaciones sólo de MD se realizaron durante 15 ns frente a un total de 100 ns de simulaciones de MD en el protocolo de GCMC-MD. Por tanto, la presente metodología logra un muestreo espacial compatible con el observado en simulaciones de MD PBC sin sesgo. Se observa que el valor de  $\mu_{ex}$  se estableció cerca de los valores de HFE en una simulación de esquema I de acetaldehído y metanol sin MD al final de cada iteración. Sin embargo, dado que las moléculas son rígidas durante GCMC, es probable que se necesiten simulaciones de MD para conservar el muestreo conformacional y espacial correcto de los solutos en estos entornos en fase en volumen.

Cálculos posteriores se centraron en determinar si el enfoque de GCMC-MD puede obtener el muestreo de soluto en equilibrio con valores de  $\mu_{ex}$  fijados conocidos. Por consiguiente, se realizó un conjunto de simulaciones de GCMC-MD, en las que en vez de empezar la simulación asignando  $\mu_{ex}$  para que sea 0, el método mantiene el valor de  $\mu_{ex}$  fijado a la HFE a lo largo de todas las iteraciones. En la figura 3 se muestran la concentración y  $\mu_{ex}$  para el sistema de acetaldehído en función de las iteraciones de GCMC-MD. En estas simulaciones, se encontró que, a medida que aumentaba el número de iteraciones y las aguas alcanzaban concentraciones en volumen, el número de intercambios de soluto disminuía de manera considerable. Se observaron tendencias similares para los demás solutos (figuras 8 y 9). Sin limitarse a una teoría específica, se cree que la disminución del número de intercambios de soluto es lo más probablemente el resultado de la búsqueda con sesgo de cavidad usada durante movimientos de GCMC. Cuando se realizan los movimientos para una molécula, las demás moléculas (tanto solutos como aguas) son estacionarias y sólo participan en interacciones no de unión con la molécula actual que está sometándose a intercambio, impidiendo por tanto movimientos solapantes de moléculas en ubicaciones ya ocupadas. Por tanto, a la concentración de agua en fase de volumen de 55 M, resulta más fácil que las moléculas de agua más pequeñas rellenen cavidades en el sistema A que los solutos más grandes, conduciendo a una disminución drástica de las probabilidades de intercambio de GCMC para los solutos (figura 3, C y D). Dado que los intercambios de GCMC continuos de fragmentos mediante inserciones, deleciones y relajación local mediante traslaciones y rotaciones son importantes para mantener el equilibrio químico entre el sistema A y los depósitos de fragmentos en fase de gas acoplados, se consideró que la disminución del número de intercambios de soluto era problemática.

Para superar esto, se hizo fluctuar de manera cíclica el  $\mu_{ex}$  tanto de los solutos como del agua alrededor de sus HFE respectivas. Se añadió y se restó de manera alternativa  $d\mu_p = \frac{N_p}{N_{p,objetivo}} a_{\mu_p}$  ( $P=F_1, F_2, \dots, F_M, W$ ) a lo largo de cada tres iteraciones de GCMC. Tales variaciones de  $\mu_{ex}$  conducen a mejoras en el muestreo a lo largo del transcurso de las iteraciones de GCMC-MD, tal como se demuestra tanto por el cambio del número de las moléculas de soluto como por el aumento de las probabilidades de movimiento de GCMC ( $P_{ins+del}$  y  $P_{tras+rot}$  para inserciones, deleciones, traslaciones y movimientos de rotación, respectivamente), tal como se muestra en la figura 3. Se observaron tendencias similares para los solutos apolares y los demás solutos polares (figura 8 y figura 9). Por tanto, junto con GCMC, tanto del disolvente como del soluto, es importante continuar haciendo fluctuar el  $\mu_{ex}$  suministrado a estas moléculas una vez que se alcanza la concentración objetivo, tal como se describe en la ecuación 2, para mantener un muestreo eficaz de los solutos en el sistema de simulación A.

## II. El sistema de mutante de lisozima L99A de T4

Se seleccionó T4-L99A, que contiene un sitio de unión ocluido modificado por ingeniería para benceno, como sistema de modelo. El método de muestreo de GCMC-MD fue idéntico a los sistemas acuosos anteriores, excepto porque el sistema B se trató como periódico con el soluto que estaba estudiándose incluido en el sistema B a la concentración objetivo usada para impulsar el muestreo de GCMC. El sistema A era una esfera de 20 Å centrada en el sitio de unión de T4-L99A definido por los residuos Ala 99 y Met 102 (figura 10).

Los cálculos del esquema I en T4 implicaron benceno como único soluto a una concentración de 1 M junto con agua a una concentración en volumen de 55 M. Estas restricciones paramétricas permitieron la validación del método, concretamente para impulsar el muestreo de benceno en la cavidad de unión ocluida de la proteína así como producir una estimación cuantitativa de unión a ligando. A lo largo de una simulación de GCMC-MD de  $10 \times 37,5$  ns con conformaciones guardadas cada 10 ps, el método mostró que el benceno se unía al sitio de unión de T4-L99A durante un total de 372 ns, estando la cavidad vacía durante tan sólo 3 ns. La concentración de benceno promedio en el sistema de simulación A fue de  $1,5 \pm 0,2$  M. Estos resultados indican que el método de GCMC-MD puede muestrear la cavidad ocluida así como alcanzar la concentración definida en todo el sistema que impulsa el muestreo de GCMC.

El muestreo de cavidad también permitió la estimación de la afinidad de unión de un ligando  $\Delta G^\circ$ , usando la siguiente ecuación:

$$\Delta G^\circ = -RT \ln \frac{[PL]}{[P]} + RT \ln [L] V_{\text{ref}} \quad (4)$$

En la ecuación (4), R es la constante de los gases (kcal/mol/K), T (K) es la temperatura, [PL] es la concentración del ligando unido, [L] es la concentración total del ligando, [P] es la concentración de la proteína,  $V_{\text{ref}}$  es el volumen de referencia en unidades de concentración ( $\sim 1660 \text{ \AA}^3$  por una molécula de ligando para 1 M de ligando). Dado que las simulaciones se mantienen en equilibrio, las concentraciones de ligando unido frente a no unido pueden correlacionarse con la fracción de tiempo de ligando unido frente a no unido en el sitio de unión a lo largo de toda la simulación. Dado que la cavidad de T4-L99A está completamente ocluida, la presencia o ausencia de un átomo de soluto en el sitio activo sólo se impulsa por los movimientos de inserción/delección de GCMC. A lo largo de la simulación de GCMC-MD de  $10 \times 37,5$  ns que implica un total de 18,75 millones de intentos de inserción/delección de GCMC para el benceno, con  $[L] \sim 1,5$  M y una razón de  $[PL]/[P]$  de 99,4/0,6,  $\Delta G^\circ$ , calculada usando la ecuación 2, es de aproximadamente -3,25 kcal/mol. Aunque hay alguna diferencia con respecto a la afinidad de unión experimental de  $-5,19 \pm 0,16$  kcal/mol, debe observarse que un aumento de  $10^4$  etapas de un total de  $1,9 \times 10^7$  etapas con benceno en la cavidad, se traduce en una diferencia de casi 1 kcal/mol de la  $\Delta G^\circ$  calculada. Esto no significa una falta de equilibrio en el sistema, sino que enfatiza la dificultad de hacer converger el cálculo de una constante de unión en un único sitio en una proteína, aunque los efectos de campo de fuerza también pueden tener un impacto sobre la  $\Delta G^\circ$  obtenida.

Los cálculos del esquema II incluyeron siete solutos junto con agua. Como con el sistema acuoso, la concentración objetivo para las disoluciones de cada soluto fue de 0,25 M. Esta simulación se realizó durante un total de  $10 \times 50$  ns. Para facilitar el análisis de los resultados, se calcularon patrones de afinidad de átomos seleccionados a partir de diferentes solutos en la cavidad de unión ocluida, denominados "FragMaps de energía libre en la cuadrícula (GFE)".

Los Fragmaps de GFE son distribuciones de probabilidad sometidas a transformación de Boltzmann para los átomos de soluto que se normalizan usando las distribuciones de moléculas de soluto en una disolución acuosa que no contiene la macromolécula. Esta normalización explica la penalización de energía libre debido a desolvatación de soluto cuando se calculan las GFE. Entonces pueden visualizarse estos mapas para evaluar cualitativamente la capacidad del método de muestreo de GCMC-MD para reproducir las posiciones de diferentes ligandos que se sabe que se unen a T4-L99A que se han sometido a análisis experimental. Se consideraron nueve ligandos: 1) benceno, 2) o-xileno, 3) p-xileno, 4) etilbenceno, 5) benzofurano, 6) indeno, 7) indol, 8) isobutilbenceno y 9) n-butilbenceno.

La figura 4 muestra los mapas de compuestos aromáticos (BENC), compuestos alifáticos (PRPC), donadores (MEOH, FORH) y aceptores (MEOO, FORO, AALO) de enlaces de hidrógeno polares junto con las orientaciones cristalográficas de los ligandos. El benceno ocupa el FragMap de compuestos aromáticos mientras que los restos alifáticos de los demás ligandos ocupan la región de FragMap de compuestos alifáticos que sobresale alejándose de la molécula de benceno. Además, los mapas de donadores y aceptores de enlaces de H polares están en la cavidad de unión y están en las inmediaciones de los grupos funcionales correspondientes en benzofurano e indol. De manera importante, aunque se usa la estructura de proteína a partir del complejo de T4-L99A-benceno (PDB 181L) como conformación de partida, las simulaciones de GCMC-MD identifican correctamente la capacidad de la cavidad para alterar su conformación para permitir una interacción favorable con los restos alifáticos así como su capacidad para adaptarse a funcionalidad polar. La capacidad del método de la invención para adaptarse a flexibilidad de proteína se debe a la aplicación de MD, que explica alteraciones en la conformación de proteína (flexibilidad de proteína) durante la unión.

El uso de Fragmaps de GFE también tiene otras ventajas. Por ejemplo, los Fragmaps de GFE permiten la evaluación cuantitativa de la afinidad relativa de ligandos, basándose en puntuaciones de GFE de ligando (LGFE), tal como se describió anteriormente.

Las puntuaciones de LGFE cuantifican el solapamiento de átomos en el ligando con los Fragmaps de GFE correspondientes. Las puntuaciones de LGFE se calcularon como promedios ponderados de Boltzmann a partir de conjuntos de orientaciones de ligando-proteína generadas usando i) muestreo de MD de los ligandos unidos a la proteína y ii) muestreo de MC de los ligandos en el campo de FragMaps. Tal como se muestra en la figura 5, tanto LGFE<sup>MD</sup> como LGFE<sup>MC</sup> se correlacionan muy bien con las energías libres de unión experimentales (alto  $R^2$  e índice de predicción (PI)). De manera importante, las LGFE pueden distinguir entre la actividad de unión de ambas series congénicas y diversas clases de ligandos. El intervalo de afinidades de unión experimentales es de -2,1 kcal/mol mientras que las puntuaciones de LGFE a partir del conjunto de MD de proteína+ligando y el muestreo de MC se extienden a lo largo de intervalos más amplios de -4,42 y -6,1 kcal/mol, respectivamente. Esto no resulta inesperado, ya que las puntuaciones de LGFE no son energías libres de unión auténticas ya que se omiten numerosos términos que contribuyen a la unión (por ejemplo, la entropía de configuración de los ligandos).

III. Mapeo de patrones de energía libre de grupos funcionales en sitios ocluidos de proteínas

Se eligieron ocho solutos representativos con diferentes funcionalidades químicas: benceno, propano, acetaldehído, metanol, formamida, imidazol, acetato y metilamonio, para examinar las LBP. El benceno y el propano sirven como

sondas para funcionalidades no polares, mientras que el metanol, la formamida, el imidazol y el acetaldehído son moléculas neutras que participan en la formación de enlaces de hidrógeno. Las moléculas de metilamonio cargada de manera positiva y de acetato cargada de manera negativa sirven como sondas para grupos donadores y aceptores cargados, respectivamente. Después se usaron las distribuciones de probabilidad normalizada para átomos seleccionados en estos solutos a partir de las simulaciones de SILCS-GCMC/MD para crear FragMaps de afinidad de grupos funcionales en las LBP respectivas tal como se describe a continuación.

#### a. Receptor de andrógenos

Para crear FragMaps, se realizó una SILCS-GCMC/MD con la estructura cristalina de testosterona-AR (PDB 2AM9) después de la retirada de testosterona. Aunque se iniciaron las simulaciones con la conformación unida a esteroide de AR, los FragMaps a partir de las simulaciones de GCMC/MD de 10x100 ns resumen las ubicaciones de diferentes grupos funcionales para ligandos cristalográficos tanto esteroideos como no esteroideos (figura 13). Por tanto, se muestra que los anillos de cicloalcano de los ligandos esteroideos que ocupan la cavidad en gran medida hidrófoba del AR tienen un buen solapamiento con los FragMaps de compuestos apolares (APOLAR) (A1 en la figura 13A). Por consiguiente, los grupos cetona de TES y dihidrotestosterona (DHT) que forman enlaces de hidrógeno con R752 se solapan con los FragMaps de compuestos cargados de manera negativa (NEG) (N1 en la figura 13A). De manera similar, los FragMaps de donadores de enlaces de hidrógeno (HBDON) y de compuestos cargados de manera positiva (POS) cerca de N705 se solapan con los grupos 17 $\beta$  de los esteroides.

De manera notable, la cavidad apolar entre W741, L873 y T877 que se ocupa por algunos ligandos tales como EM-5744 esteroideo y S-1 no esteroideo, un análogo de R-bicalutamida, es inaccesible en la conformación de partida. Sin embargo, tras la simulación de la estructura, las cadenas laterales de W741 y T877 experimentan cambios conformacionales que conducen a la formación de una cavidad sin afectar significativamente a la conformación global del receptor de andrógenos. Por ejemplo, la desviación de media cuadrática (RMSD) en la estructura principal de proteínas es de ~1,2 Å, por consiguiente, se encontraron FragMaps de APOLAR en esta cavidad (A2, figura 13A). Tal como se ilustra en la figura 13D, no se encontraron densidades de FragMaps de APOLAR en una simulación sólo de GCMC en la que la proteína era rígida. Esta comparación valida que la flexibilidad de proteína mediante la inclusión de MD es importante dado que permite que los solutos muestreen regiones de la proteína que no estaban disponibles en la conformación de partida. Una evidencia adicional de la importancia de incluir flexibilidad de proteína fue el notable aumento en la zona muestreada mediante FragMaps de APOLAR en el sitio A1 en la simulación de GCMC/MD frente a sólo GCMC (figuras 13A frente a 13D). Los inventores creen que este aumento del muestreo está impulsado por las flexibilidades de las cadenas laterales de residuos que forman la cavidad. Tomados en conjunto, estos resultados apuntan a la capacidad cualitativa del enfoque de SILCS-GCMC/MD para mapear los requisitos de grupos funcionales de una LBP completamente ocluida, incluyendo la capacidad del método de SILCS para identificar regiones accesibles para solutos significativamente más allá de las presentes en la estructura cristalina.

#### b. PPAR $\gamma$

Para mapear los patrones de energía libre de grupos funcionales en las LBP de PPAR $\gamma$  se inició un SILCS-GCMC/MD 10x50 ns usando la estructura 3U9Q de PDB después de la retirada del ácido decanoico de la estructura cristalina. Tal como se ilustra en la figura 14A, los FragMaps muestrean la cavidad del ácido decanoico (marcada LBP1 en la figura 14A), y una segunda cavidad que está flanqueada entre las hélices H3, H4 y las láminas  $\beta$  B2 y B3 (LBP2). Entre las dos cavidades, la cavidad LBP1 está más ocluida que la cavidad LBP2 (véase la tabla 6) y, por tanto, se produce cierto muestreo de LBP2 incluso con la conformación rígida de proteínas en las simulaciones sólo de GCMC. Sin embargo, hay un aumento en el grado del muestreo de FragMaps cuando la flexibilidad de la proteína se incluye a través de MD.

Tabla 6. Capacidad de oclusión de una cavidad de unión de ligando (LBP) medida como una razón del área superficial accesible al disolvente (SASA) de un ligando en volumen frente a la LBP.

Receptor	Ligando	SASA en volumen	SASA de LBP	Razón
AR	Testosterona	402,1	0	0
PPAR $\gamma$	Rosiglitazona	506,8	68,6	0,13
	Ácido decanoico	341,1	28,8	0,08
mGluR	FITM	576,8	43,8	0,07
$\beta_2$ AR	BI-167107	534,8	36,32	0,07
	Carazalol	478,6	62,8	0,13

Además, tal como se muestra en la figura 15, los FragMaps trazan una trayectoria desde la superficie de la proteína hasta las LBP, lo que indica una posible trayectoria para la unión del ligando.

La validación adicional del método se proporciona por un mayor grado de solapamiento entre la cadena de alquilo terminal del ácido decanoico y las densidades de FragMap de APOLAR, así como el solapamiento entre el ácido carboxílico del ácido decanoico y las densidades de FragMap de NEG (A1, N1, respectivamente) en la LBP1 (figura

14A). También se encontró que los diferentes grupos funcionales de rosiglitazona, un fármaco antidiabético conocido que se une a PPAR $\gamma$  en la cavidad de LBP2, se solapan bien con los FragMaps (figura 14). Por ejemplo, el resto de tiazolidindiona se solapa con los FragMaps de NEG en las proximidades de H323 y H449 (N1 en la figura 14B), mientras que el resto de piridina se solapa con los FragMaps de APOLAR (A3) en las proximidades de M364 y V339 y el grupo de unión de etoxilo entre el la tiazolidindiona y la piridina se solapan con los FragMaps de HBACC (HBA1).

Junto con estas observaciones cualitativas, se estimaron las afinidades de unión de los ligandos usando la puntuación de energía libre en la cuadrícula de ligandos (LGFE) para 16 ligandos cuyos datos de actividad de unión al PPAR $\gamma$  humano están disponibles, y se compararon con la afinidad de unión experimental,  $\Delta G_{\text{unión}}$ . Se normalizaron las  $K_i$  obtenidas de las diferentes fuentes frente a la  $K_i$  de rosiglitazona ( $K_i = 120$  nM; la DB de unión notificó un intervalo de valores de entre 8-440 nM). A pesar de la diversidad en los ligandos y sus modos de unión, existe una correlación razonable entre los valores de LGFE y  $\Delta G_{\text{unión}}$  con un índice predictivo (PI) de  $\sim 0,63$  y  $R^2 -0,22$ . Por ejemplo, el compuesto GW409544 que se une a las cavidades LBP1 y LBP2 (PDB: 1K74) tiene una solapamiento muy bueno con los FragMaps de APOLAR A1, A2 y A3 (figura 14C), lo que lleva a una LGFE favorable que se correlaciona con su alta afinidad de unión en comparación con el agonista parcial ácido decanoico, o las tiazolidindionas tales como la rosiglitazona. Por otro lado, se observan malas correlaciones para Cerco-A (figura 14D) debido a la falta de los FragMaps de APOLAR en la cavidad hidrófoba entre L262 y F287, donde se une el grupo funcional dibenzofurancaboxamida de Cerco-A. Los presentes inventores han planteado la hipótesis de que la escasa correlación se debe a la pérdida de la cavidad hidrófoba debido a 1) las altas flexibilidades de las cadenas laterales de estos residuos durante las simulaciones y 2) la conformación de las hélices de conexión de bucle H2 y H3, modeladas utilizando el programa MODELLER, tal como se describió anteriormente. Aunque algunas de las cadenas laterales son flexibles, sin embargo se conserva la conformación general del receptor. De hecho, la LGFE de ligandos calculada usando FragMaps a partir de las simulaciones sólo de GCMC se correlacionan pero con el  $\Delta G_{\text{unión}}$  experimental. Dado que los procedimientos de SBDD normalmente examinan ligandos basándose en una estimación de sus afinidades de unión, este resultado valida adicionalmente la importancia de incorporar flexibilidad de proteínas en tales estudios.

#### c. mGluR:

Para muestrear eficazmente la LBP parcialmente ocluida de mGluR1, se realizó SILCS-GCMC/MD de 10x50 ns en un monómero de mGluR obtenido de una estructura cristalina de la conformación inactiva de la región transmembrana (TM) 7 dimerica del mGluR1 de la familia C en complejo con FITM, un modulador alostérico negativo (NAM). Durante este muestreo, el GCMC de solutos y agua se restringió a una región cúbica de 20 Å alrededor de las LBP. Los FragMaps resumen correctamente los diferentes grupos funcionales del FITM (figura 16). La naturaleza en gran medida hidrófoba de la cavidad se traza por las altas afinidades de los FragMaps de APOLAR. Dos densidades de APOLAR distintas se solapan bien con la amina de pirimidina y los restos de p-fluorofenilo de FITM, que se encuentran en las proximidades de V753, V664 y 1812 (A1 en la figura 16A) y F801, 1797, W798 y L757 (A2), respectivamente. Los Fragmaps de HBDon en las proximidades de T815, que se ha observado que es importante para la unión, se solapan bien con la pirimidina de FITM. Aunque se encuentra que algunas de estas cadenas laterales son flexibles a través de la simulación de GCMC/MD, se conserva la conformación general del receptor, con la cavidad de unión estrecha.

Los estudios de mutaciones del mGluR1 y la actividad de unión con otros moduladores alostéricos como la 2-metil-6-(feniletinil)piridina (MPEP) (otro NAM conocido) revelaron que la cavidad de unión identificada con FITM podría compartirse con otros NAM. El muestreo de MC de MPEP en la cavidad y en presencia de FragMaps tal como se describió anteriormente produce modos de unión similares a los de FITM (figura 16). Las LGFE para FITM, MPEP y otros análogos de FITM se correlacionaron bien con sus  $\Delta G_{\text{unión}}$  (SI figura 17C). Aunque algunos muestreos de la cavidad se realizan a través de una simulación sólo de GCMC (figura 17A frente a 17B), de manera similar a PPAR $\gamma$ , las correlaciones fueron malas cuando se calcularon las puntuaciones de LGFE usando FragMaps a partir de una simulación sólo de GCMCA de mGluR (figura S17C frente a 17D).

#### d. $\beta_2$ AR:

Están disponibles las estructuras cristalinas de las conformaciones inactiva y activa de la familia A de GPCR y se realizaron dos SILCS-GCMC/MD de 10x50 ns independientes con cada estructura. En el presente documento, B2I se refiere a las simulaciones que comienzan con la conformación inactiva (PDB: 2RH1) y B2A se refiere a las simulaciones que comienzan con la conformación activa (PDB: 3P0G).

Se obtuvieron buenos solapamientos entre los FragMaps y los ligandos cristalográficos tanto para B2I como para B2A. Las afinidades de FragMap de APOLAR cerca de la región hidrófoba definida por F289, V117 y A200 se solapan bien con los restos de benzoxazina y carbazol de B1161707 y con carazolol, respectivamente (A1 en la figura 12). Tanto los FragMaps de POS como de HBDon adyacentes a D113 se solaparon con los grupos funcionales de amina en ambos ligandos (DN2 y D2 en la figura 12A). Aunque los FragMaps resumen la ubicación de los ligandos respectivos para B2I y B2A, se encontraron diferencias claras entre los FragMaps para estas dos conformaciones.

Se encontró una segunda densidad de APOLAR cerca de la segunda cavidad hidrófoba adyacente a W109, 1309 y F193 sólo en B2A (A2, figura 12A). La falta de esta cavidad hidrófoba en B2I se debe a la flexibilidad relativamente alta de las cadenas laterales 1309 y F193 en B2I en comparación con B2A (SI figura 18A). A pesar de estas flexibilidades de cadena lateral, se mantuvieron las conformaciones generales activa e inactiva de  $\beta_2$ AR (Fig. S18B). Los estudios mutacionales han demostrado que las interacciones agonistas con 1309 son importantes tanto para la selectividad de  $\beta_2$ AR como para la activación del ligando. Los FragMaps de HBDON y HBACC resumieron correctamente importantes interacciones entre el oxígeno del carbonilo, los grupos amina e hidroxilo de BI-167101 con S203, S204 y S207 en B2A, mientras que el nitrógeno polar solitario del heterociclo carbazol en carazolol se resumió por mapas de HBDON más estrechos en B2I (HBD1, figura 12A). Estas diferencias pueden atribuirse a las mayores flexibilidades de las cadenas laterales S203 y S204 en B2I (SI figura S18A). En consonancia con estos FragMaps diferenciales, existen estudios mutacionales que han identificado que S203 y el S204 son importantes para la unión del agonista y la activación del receptor a través de sus hidroxilos de catecolaminas.

En particular, el enfoque de SILCS-GCMC/MD también es capaz de diferenciar cuantitativamente entre los dos estados de  $\beta_2$ AR. Se obtuvieron puntuaciones de LGFE a partir del muestreo de MC para una amplia variedad de agonistas, agonistas parciales y antagonistas/agonistas inversos. Tal como se muestra en la tabla 7, se obtuvieron buenas correlaciones entre las puntuaciones de LGFE y las afinidades de unión de los agonistas y los agonistas parciales con B2A ( $R^2 \sim 0,46$ ,  $PI \sim 0,59$ , figura 12B-(1)) mientras que el mismo conjunto de ligandos produjo correlaciones significativamente peores con los FragMaps de B2I ( $R^2 \sim 0,10$ ,  $PI \sim 0,31$ , figura 12B-(2)). De manera similar, las afinidades de unión de los antagonistas/agonistas inversos se correlacionaron bien con las LGFE puntuadas usando FragMaps de B2I ( $R^2 \sim 0,45$ ,  $PI \sim 0,67$ , figura 12B-(4)), mientras que se encontró una peor correlación con las LGFE calculadas usando los FragMaps de B2A ( $R^2 \sim 0,11$ ,  $PI \sim 0,38$ , figura 12B-(3)). En consonancia con la calidad de las correlaciones, el muestreo de MC de ligandos produjo modos de unión similares a las orientaciones cristalográficas de BI167107 y carazolol (figuras 19 y 20).

Examen de ligandos guiado por FragMaps para  $\beta_2$ AR:

Los  $\beta_2$ AR se expresan en numerosos tipos de células, incluyendo el músculo liso de las vías respiratorias (ASM). La activación de  $\beta_2$ AR en las ASM produce broncodilatación y los beta-agonistas inhalados son la terapia principal para el asma y otras enfermedades pulmonares obstructivas. Se usaron las diferencias en los FragMaps entre los dos estados finales para guiar estudios de examen virtual (VS) para identificar nuevos agonistas para  $\beta_2$ AR. Siguiendo el procedimiento descrito anteriormente, se seleccionaron 15 ligandos químicamente diversos de mayor puntuación (figura 21) de una base de datos *in silico* que contiene aproximadamente 1,8 millones de compuestos. Se estudió el efecto de los compuestos seleccionados a través de una respuesta de relajación de los anillos traqueales de muestras de pulmón de ratones. Este método *ex vivo* es una estrategia de rendimiento relativamente alto y un factor de predicción mejor de la disposición macromolecular *in vivo* que los estudios *in vitro*. Se usó isoproterenol como control positivo. Siete de estos ligandos efectuaron diversos grados de relajación traqueal (figura 12C), lo que representa una tasa de éxito del 46%. Las conformaciones acopladas de los ligandos seleccionados tuvieron buenos solapamientos con los FragMaps de B2A lo que condujo a altas puntuaciones de LGFE (figura 22). Todos los ligandos ocuparon la cavidad hidrófoba definida por los residuos 1309, W109 y F193 y mantuvieron interacciones con Asp 113. También se observó que los ligandos que producían relajación traqueal mantenían interacciones con S203 y S207 en la hélice H5 identificada como importante para la activación del ligando y la selectividad de  $\beta_2$ AR. Estos resultados apuntan a la utilidad de la metodología de SILCS-GCMC/MD en el diseño racional de ligandos, incluyendo la identificación de agonistas.

Tabla 7

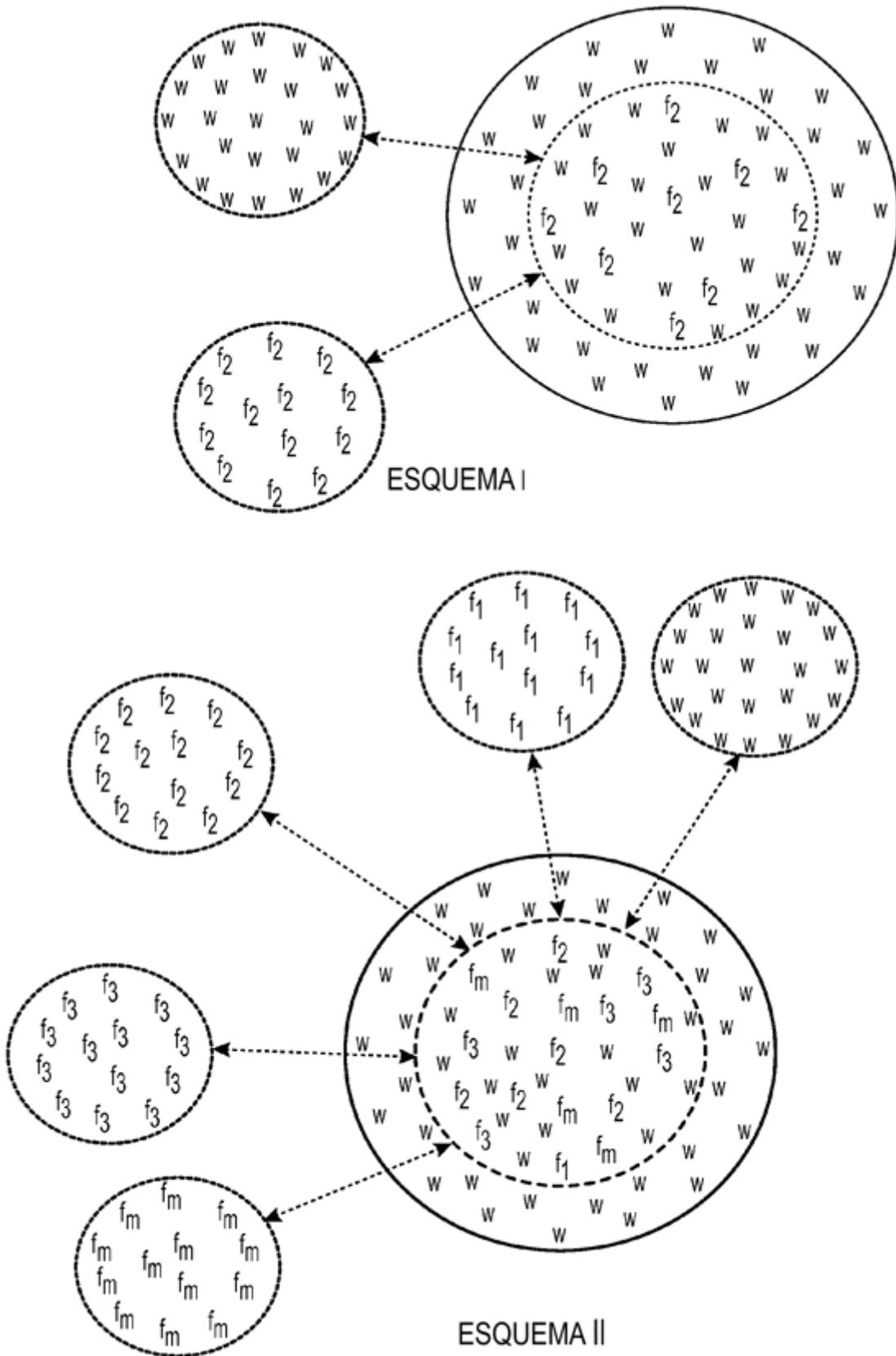
	PDB	Compuesto	Función	$\Delta G_{unión}$	Referencia
1		Salbutamol	Agonista / agonista parcial	-8,09	46
2	4LDO	Adrenalina (epinefrina)		-8,60	47
3		Isoproterenol		-9,63	47
4	4LDL	Hidroxibencil-isoproterenol		-9,72	48
5		Indacaterol		-10,19	49
6		Formoterol		-10,60	49
7		THR-144877(AA1)		-11,04	49
8		Picumeterol		-12,23	49
9		Salmeterol		-12,00	50
10	3P0G	BI-167107		-13,69	21
11		Practolol	Antagonistas/agonista inverso	-6,77	50
12		Atenolol		-8,14	50
13		Acebutalol		-8,26	50
14		Bisoprolol		-9,10	50
15		Labetolol		-10,91	50
16		Propranolol		-12,34	50

ES 2 808 724 T3

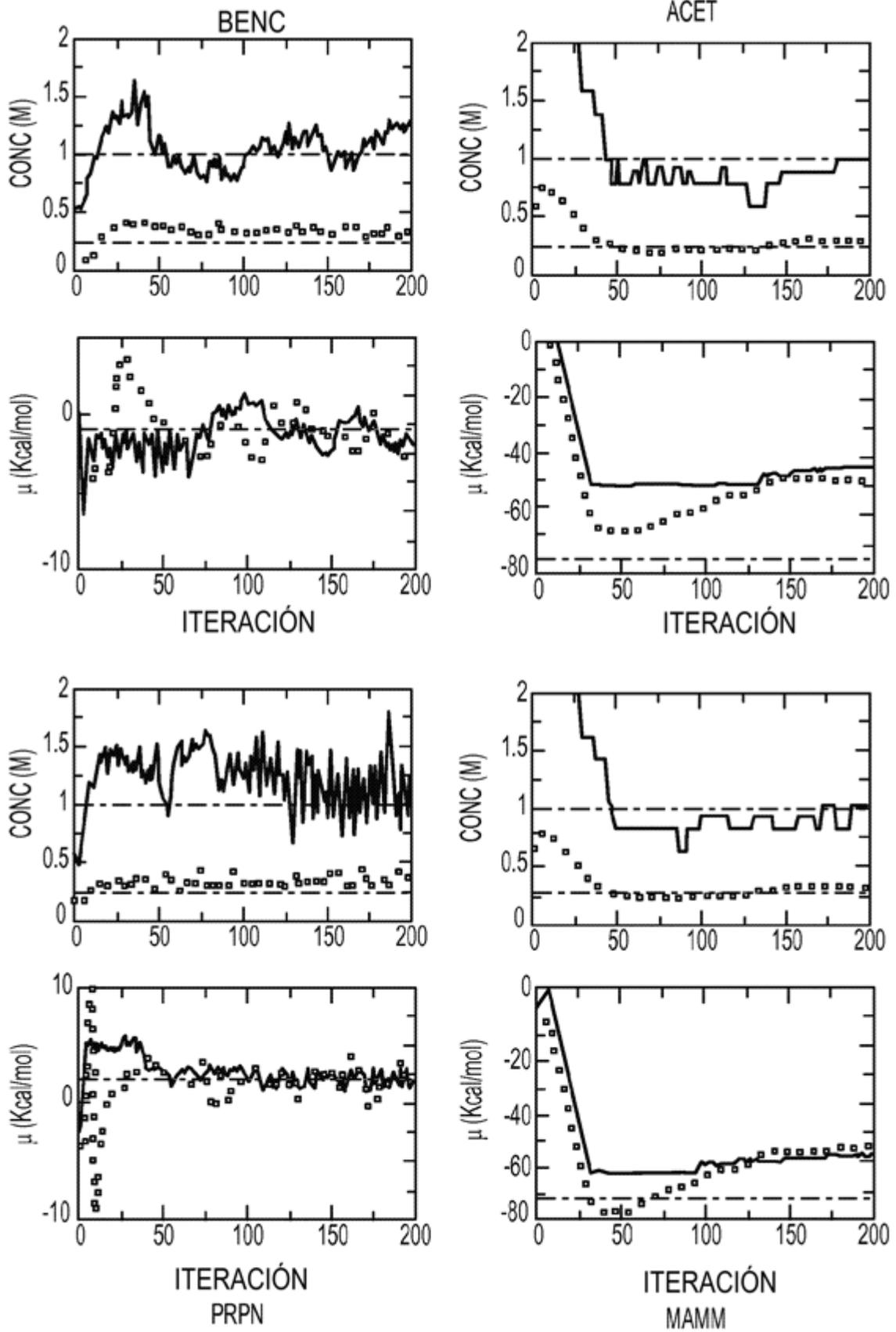
17		Pindolol		-12,43	51
18		ICI-118551		-12,58	50
19		Carvedilol		-12,77	50
20	3D4S	Timolol		-13,15	50
21	2RH1	Carazolol		-13,35	52

## REIVINDICACIONES

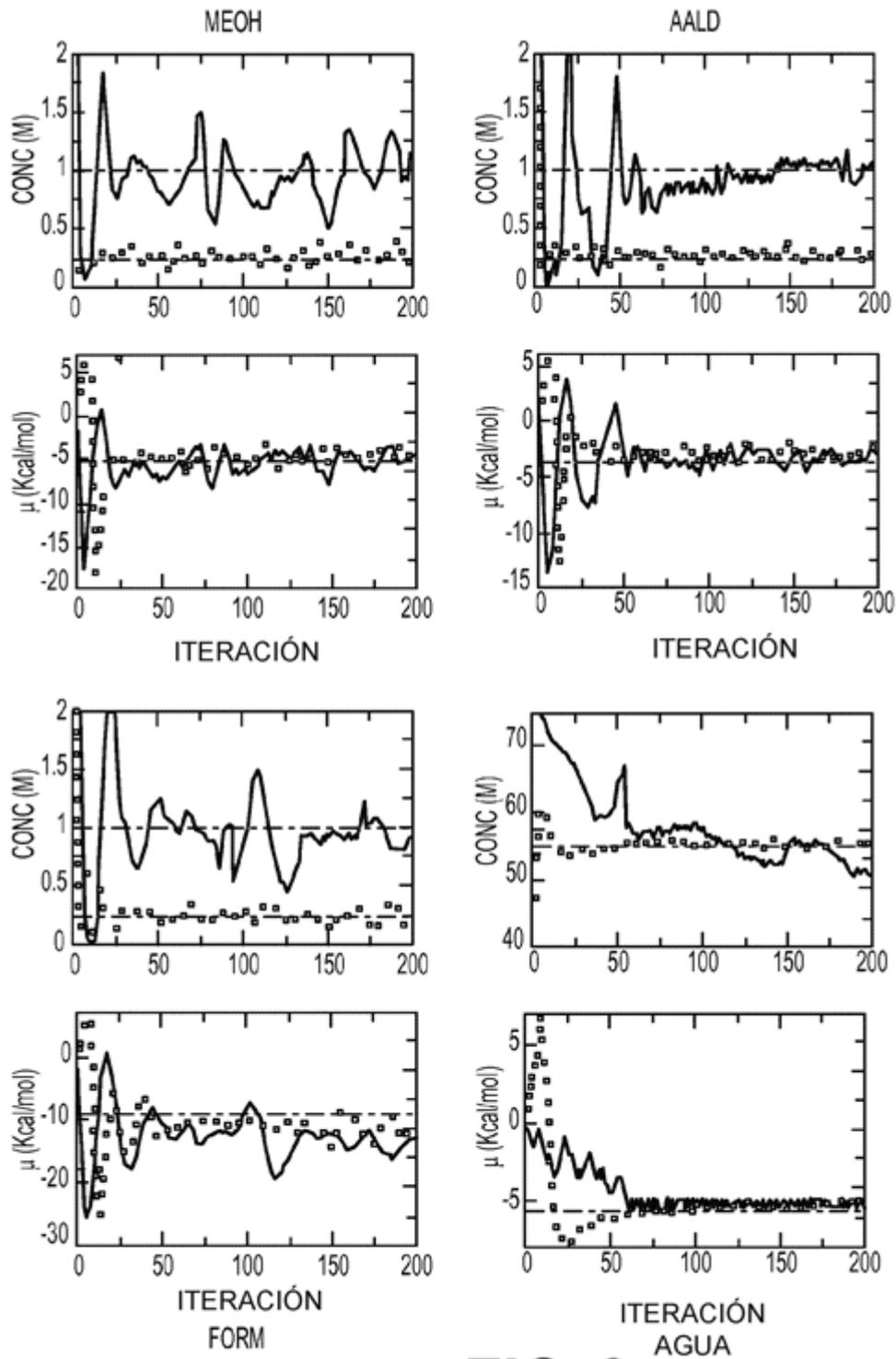
- 5 1. Método computacional para muestrear la distribución espacial de uno o más solutos y agua en una región del espacio definida (sistema) que comprende:
- 1) asignar una concentración objetivo,  $N_{obj}$ , de cada uno del uno o más solutos y agua;
- 2) muestrear la distribución espacial del uno o más solutos y agua en una región del espacio definida computacionalmente usando criterios de muestreo de Metropolis de Montecarlo de conjunto macrocanónico (GCMC), en el que el exceso de potencial químico ( $\mu_{ex}$ ) asignado para cada uno del uno o más solutos, si está presente, y agua se ajusta a 0, si está presente;
- 3) actualizar  $\mu_{ex}$  de cada uno del uno o más solutos y agua a partir de la diferencia en la concentración actual en la región del espacio definida ( $N_{sis}$ ) y el objetivo ( $N_{obj}$ );
- 15 4) repetir las etapas 2) y 3) usando los valores actualizados de  $\mu_{ex}$  en la etapa 2) para obtener una distribución espacial del uno o más solutos y agua;
- 20 en el que el sistema comprende además una o más macromoléculas, en el que la distribución espacial de uno o más solutos o agua se usa para identificar la afinidad preferente de cada uno de los solutos o agua por cada una de la una o más macromoléculas, y en el que dicha una o más macromoléculas se seleccionan de una proteína, ARN, ADN, hidrato de carbono, lípido, compuesto químico orgánico, compuesto químico inorgánico o una combinación de los mismos.
- 25 2. Método según la reivindicación 1, en el que el sistema contiene agua y un soluto y en el que la concentración objetivo del soluto y agua se ajusta a 1 M y 55 M, respectivamente, a partir de lo cual se obtiene la energía libre de hidratación del soluto a partir del valor de  $\mu_{ex}$
- 30 3. Método según la reivindicación 1, en el que el  $\mu_{ex}$  de uno o más solutos y agua se aumenta y se disminuye alternativamente durante las operaciones de GCMC a través de ciclos consecutivos que implican las etapas 2)-4), una vez que la concentración de los solutos y el agua alcanzan su valor objetivo.
- 35 4. Método según la reivindicación 1, en el que la proporción de etapas de GCMC para cada uno del uno o más solutos y agua se asigna basándose en la concentración objetivo de cada uno de los solutos y agua.
5. Método según la reivindicación 1, en el que el sistema que contiene los solutos y/o agua está englobado en un sistema más grande que contiene dichos solutos y/o agua.
- 40 6. Método según la reivindicación 5, en el que el sistema incluye adicionalmente una o más macromoléculas.
7. Método según la reivindicación 1, en el que se muestrea la distribución espacial que sigue a la etapa 2) con una simulación de dinámica molecular.
- 45 8. Método según la reivindicación 1, en el que  $\mu_{ex}$  se aumenta o se disminuye igual a  $N_{obj}/N_{sis}$ .
9. Método según la reivindicación 1, en el que  $\mu_{ex}$  se aumenta cuando  $N_{sis}$  es menor que  $N_{obj}$  y se disminuye cuando  $N_{sis}$  es mayor que  $N_{obj}$ .
- 50 10. Método según la reivindicación 1, en el que dicha región definida computacionalmente comprende una región interior que contiene el uno o más solutos y agua, ubicada dentro de una región exterior más grande que contiene agua adicional.
- 55 11. Método según la reivindicación 1, en el que se genera una salida de las distribuciones espaciales del uno o más solutos y agua.



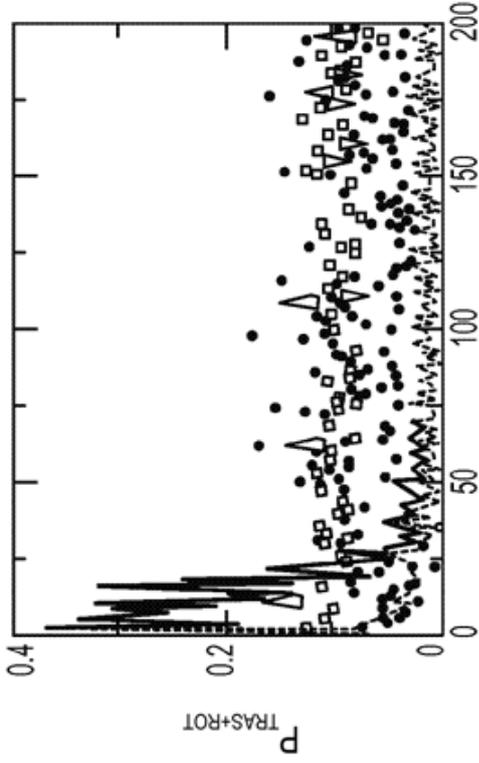
**FIG. 1**



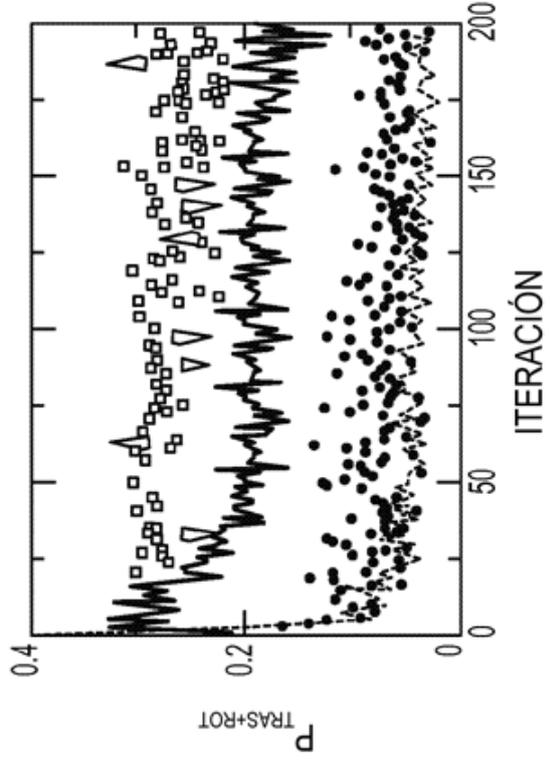
**FIG. 2**



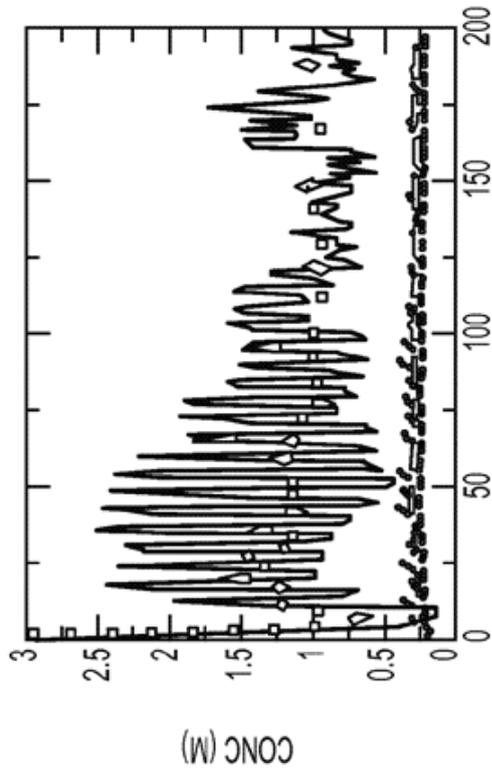
**FIG. 2**  
**CONT.**



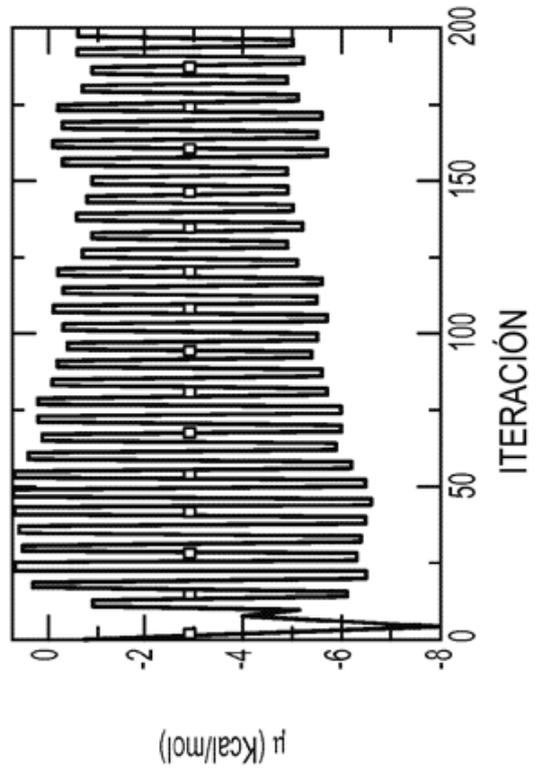
**FIG. 3C**



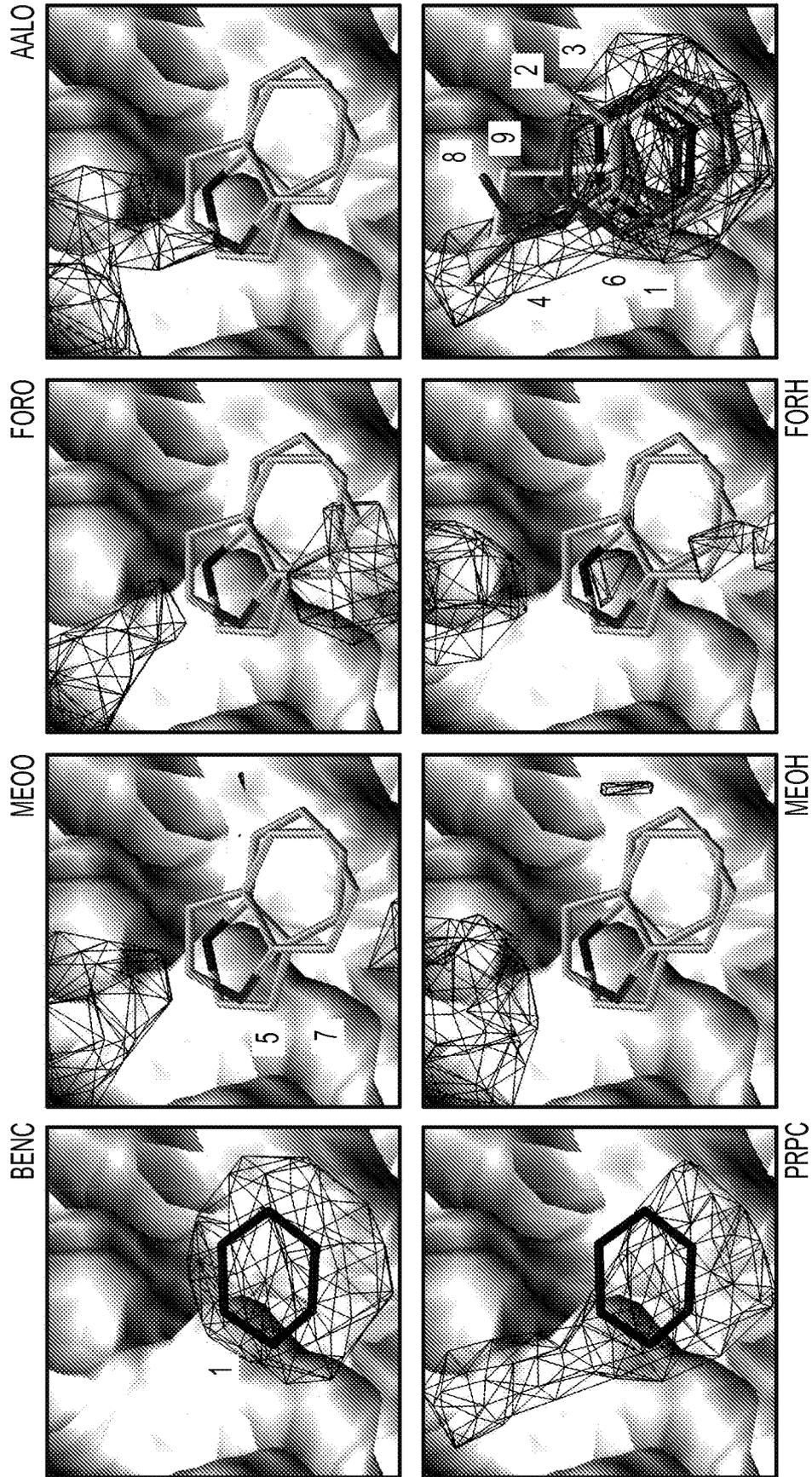
**FIG. 3D**



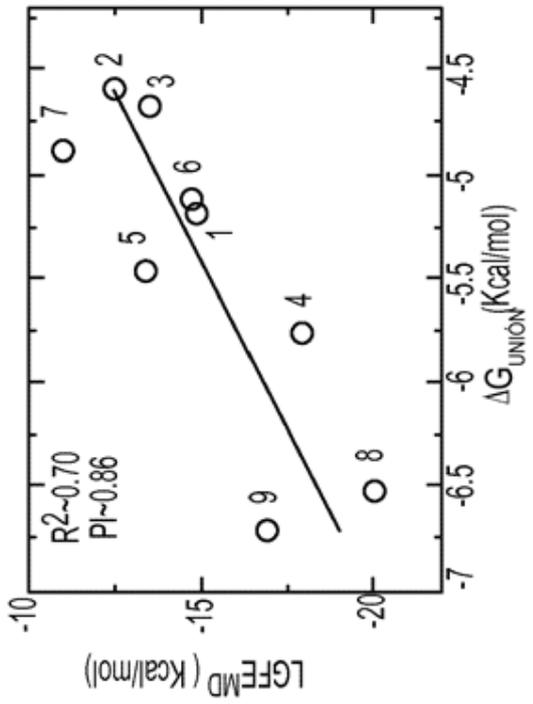
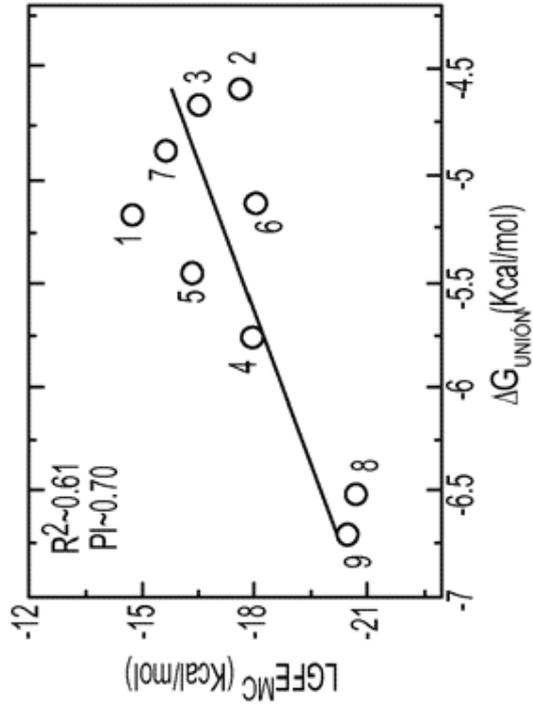
**FIG. 3A**



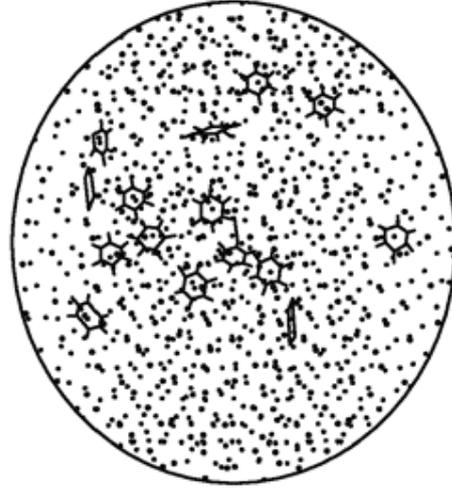
**FIG. 3B**



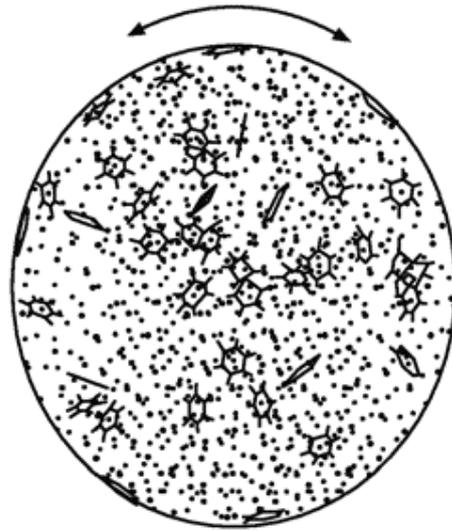
**FIG. 4**



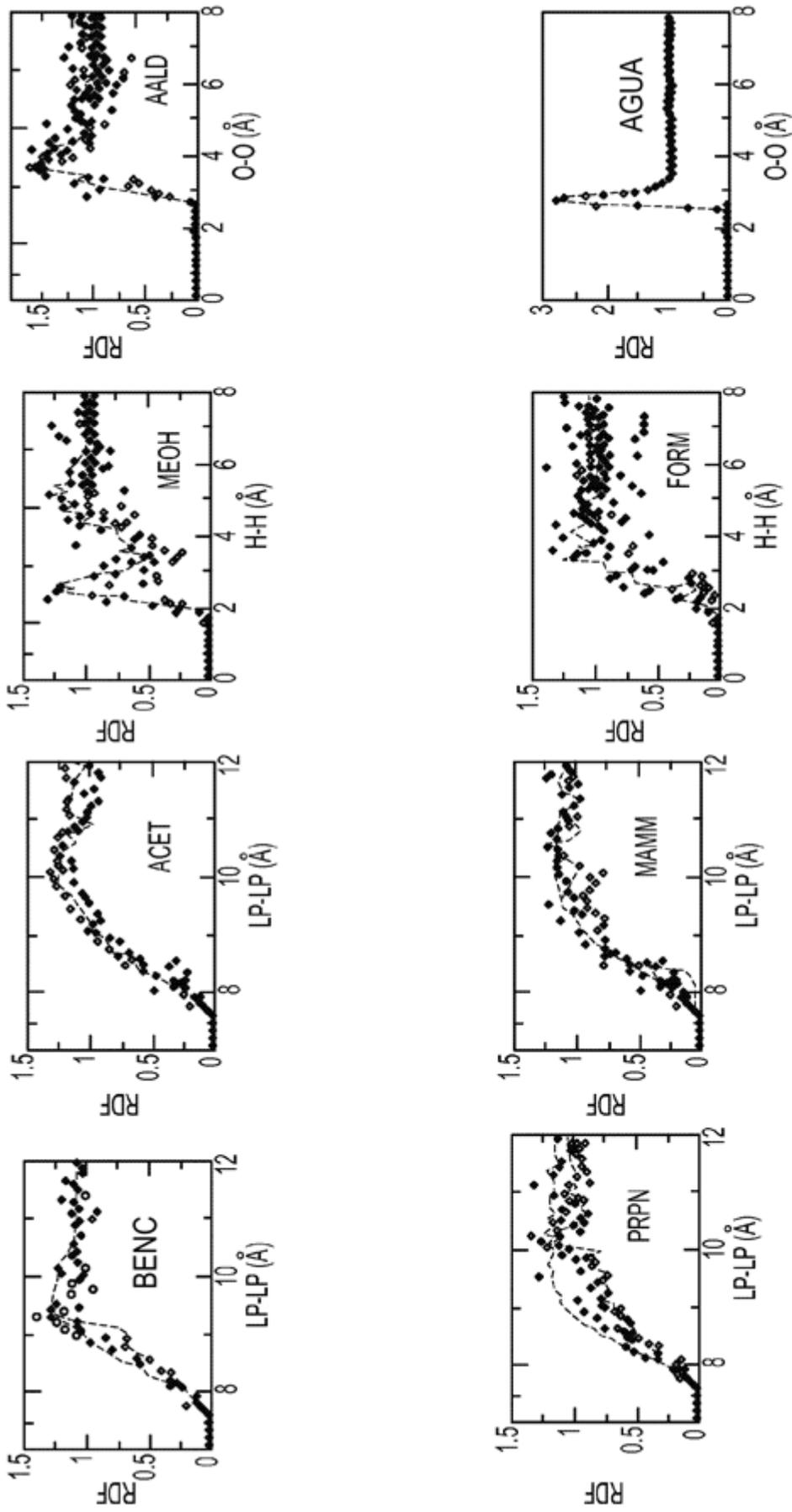
**FIG. 5**



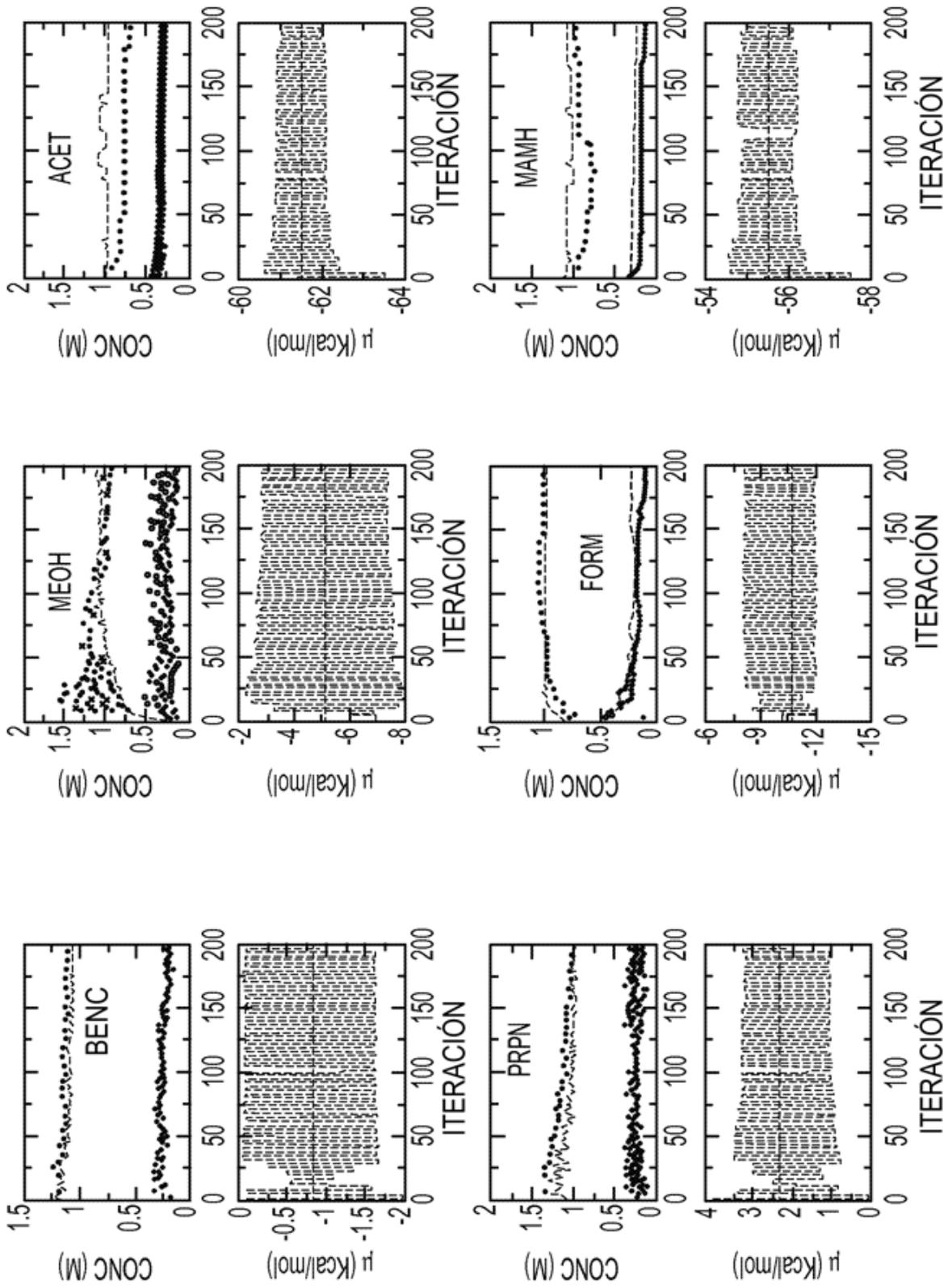
**FIG. 6B**



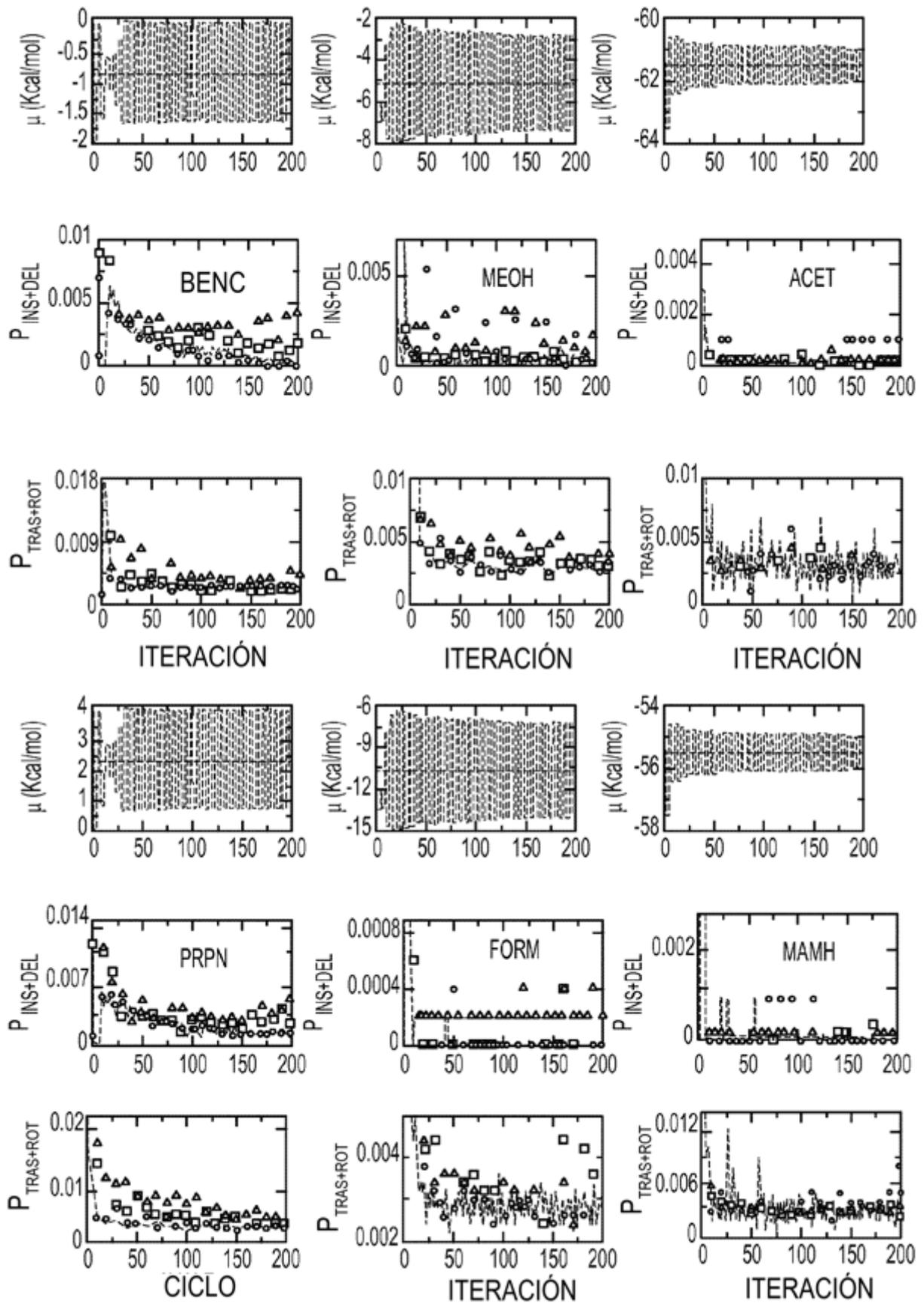
**FIG. 6A**



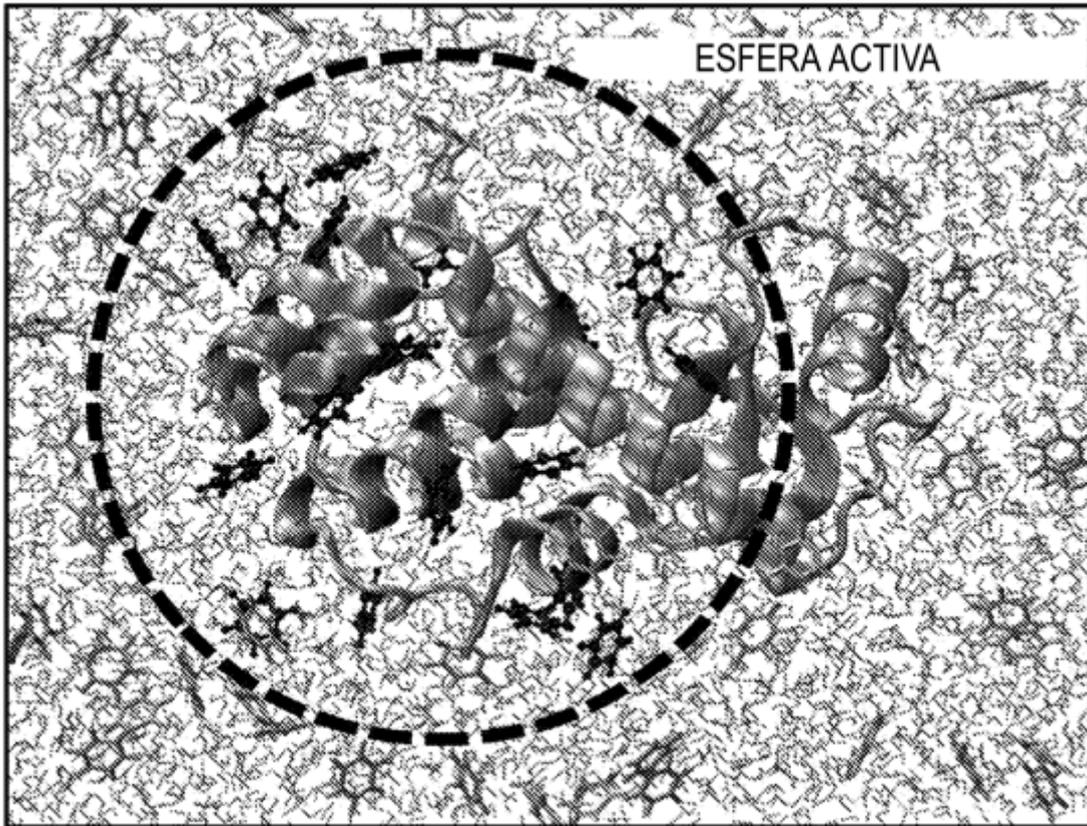
**FIG. 7**



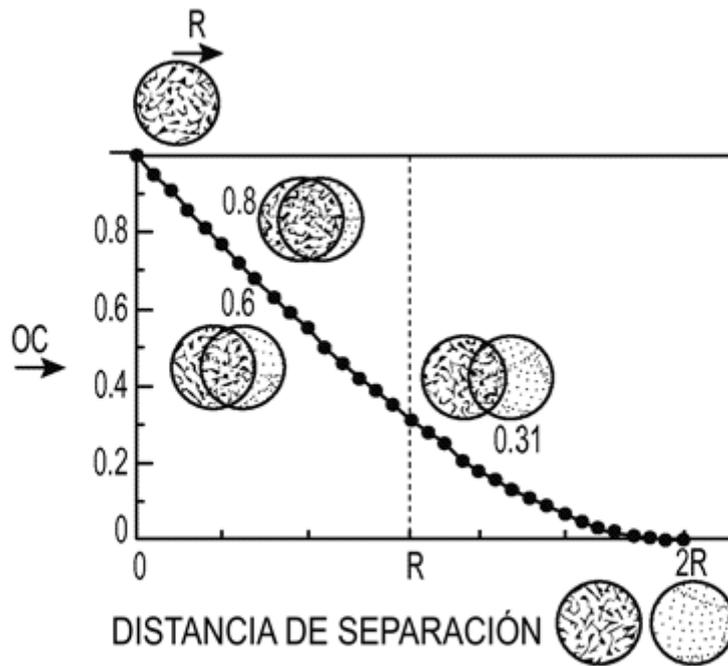
**FIG. 8**



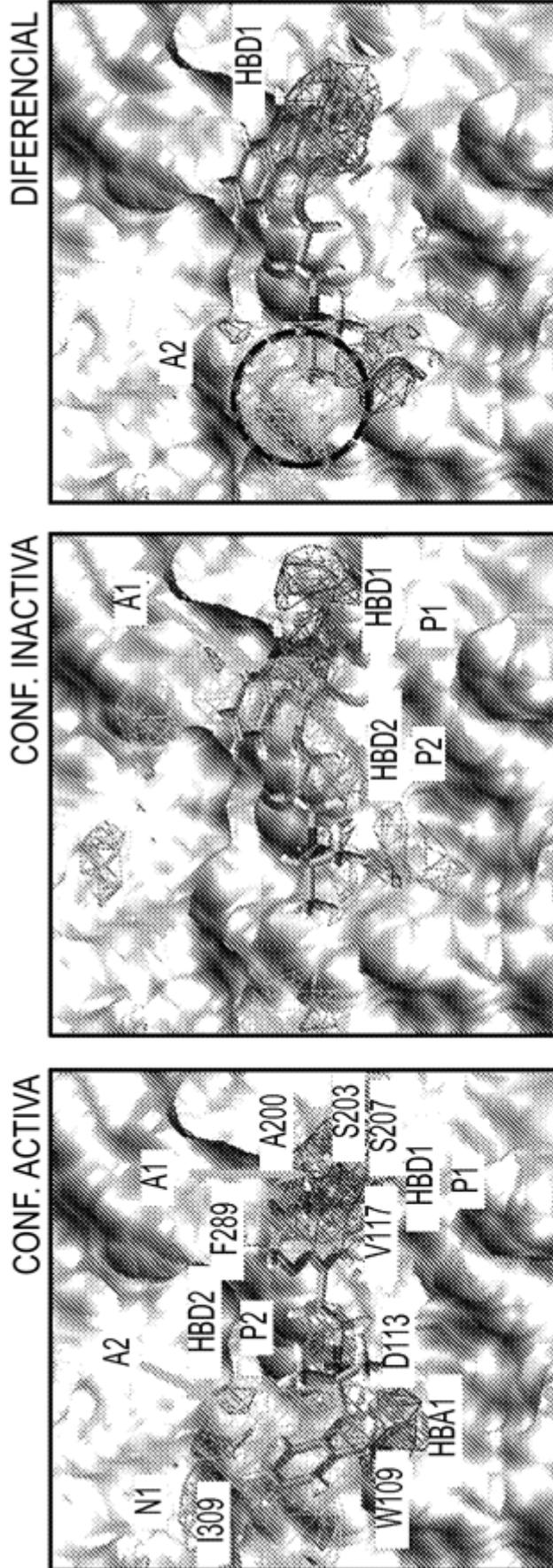
**FIG. 9**



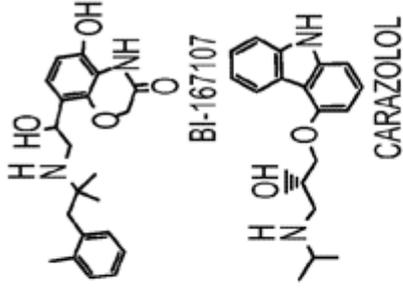
**FIG. 10**



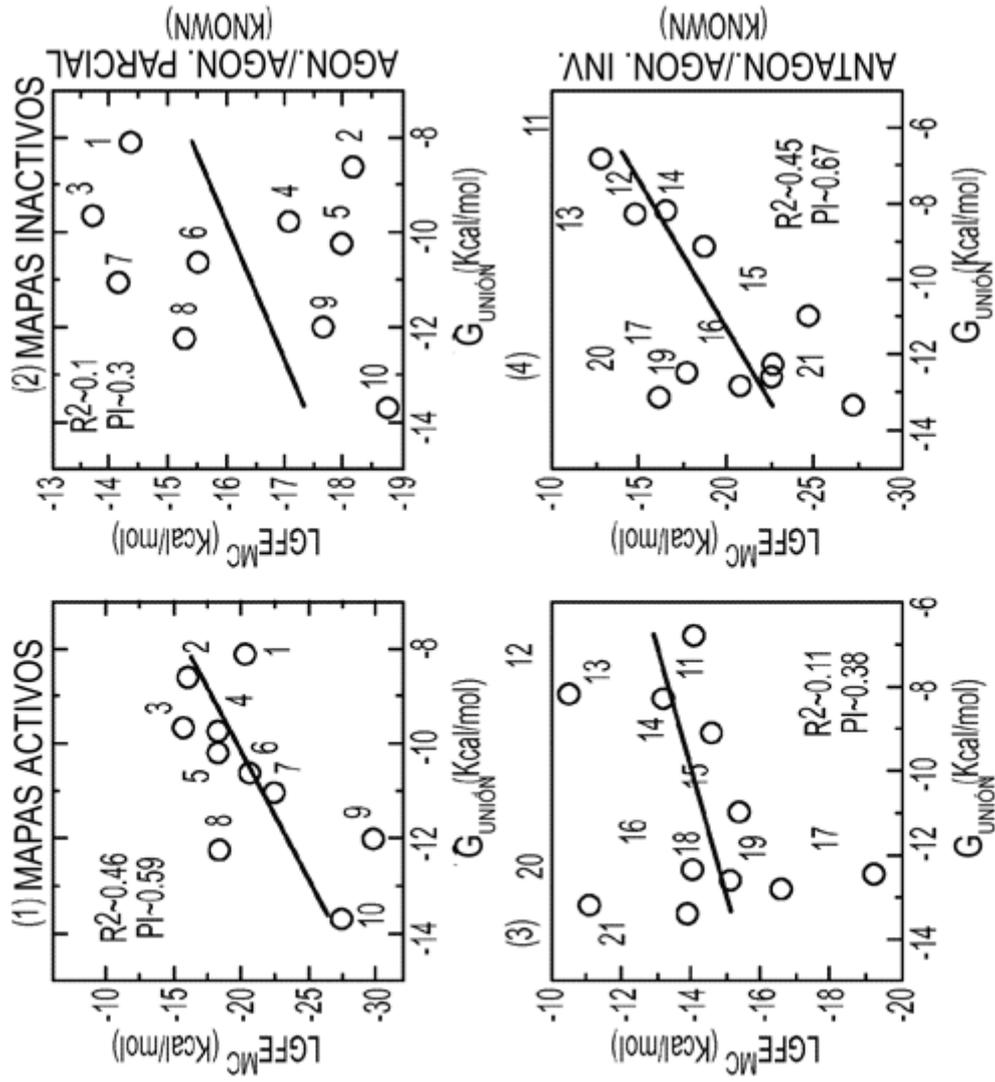
**FIG. 11**



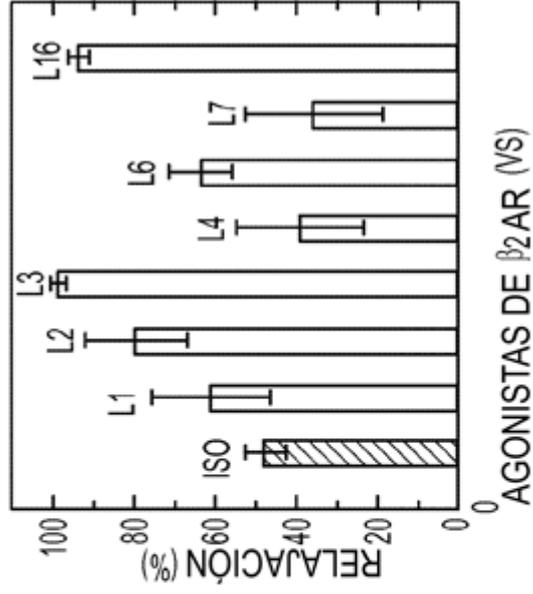
**FIG. 12A**



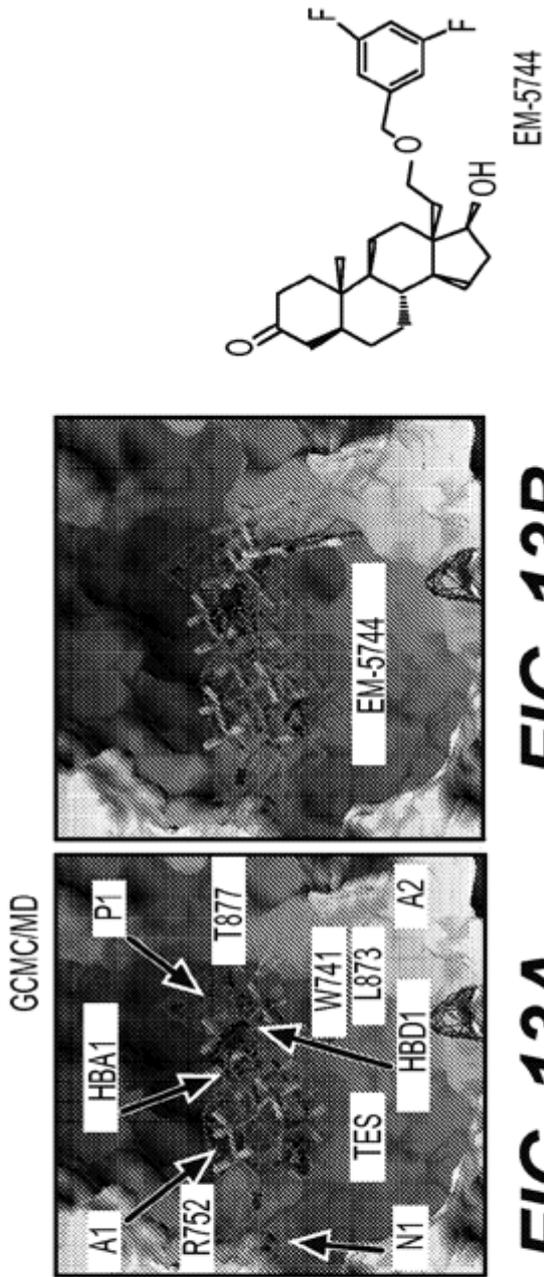
**FIG. 12C**



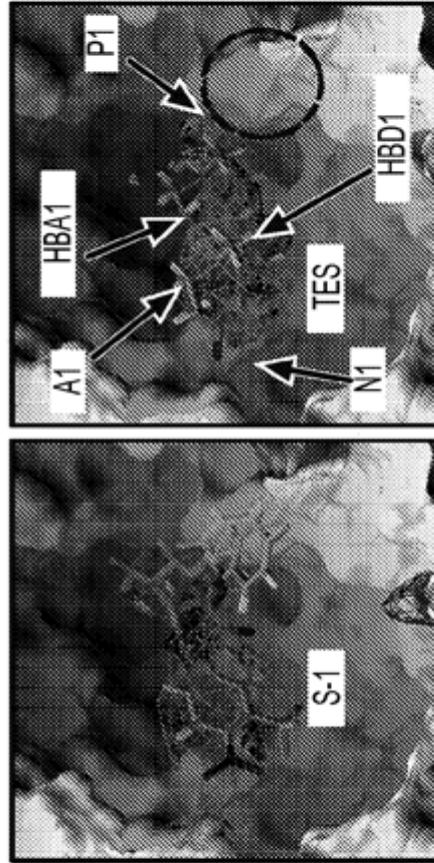
**FIG. 12B**



**FIG. 12D**

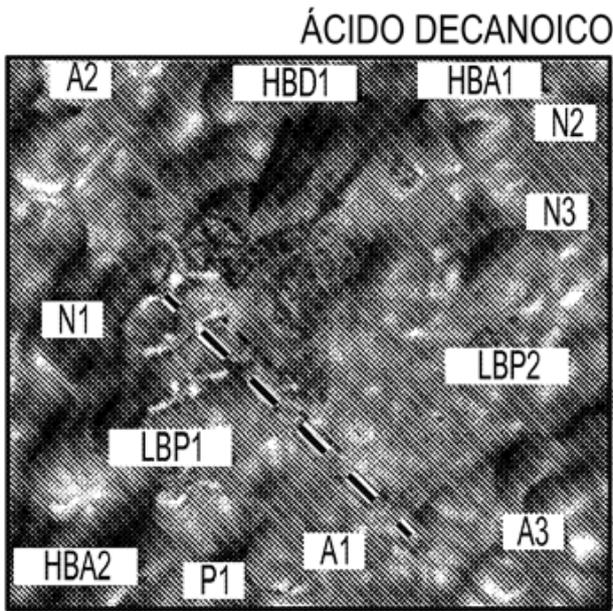


**FIG. 13B**

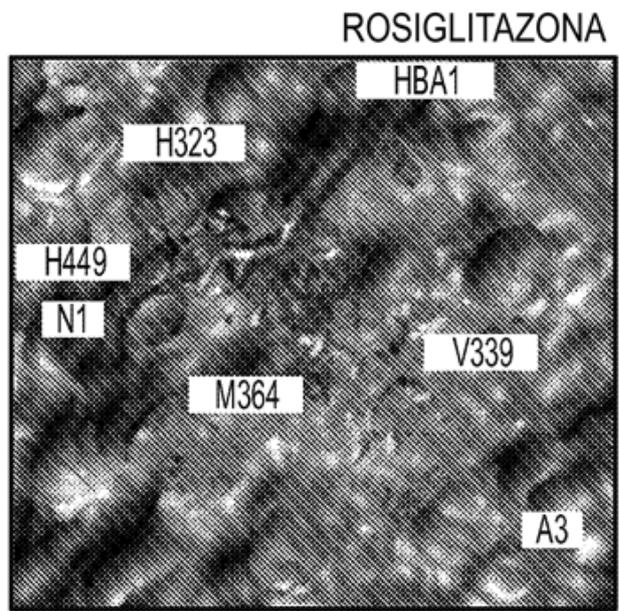


SÓLO DE GCMC

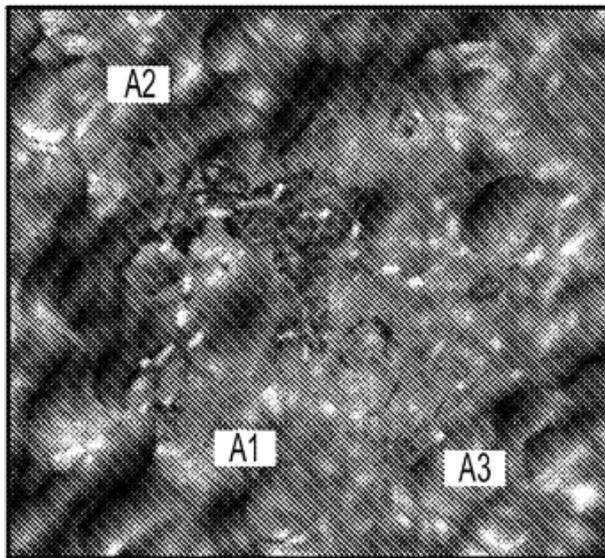
**FIG. 13D**



**FIG. 14A**

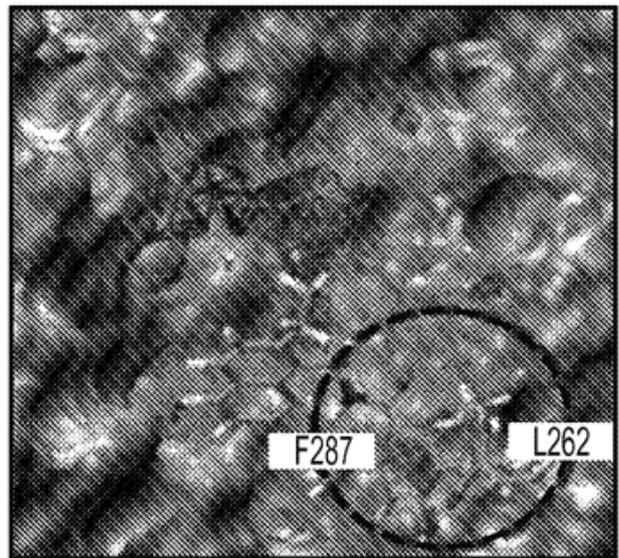


**FIG. 14B**



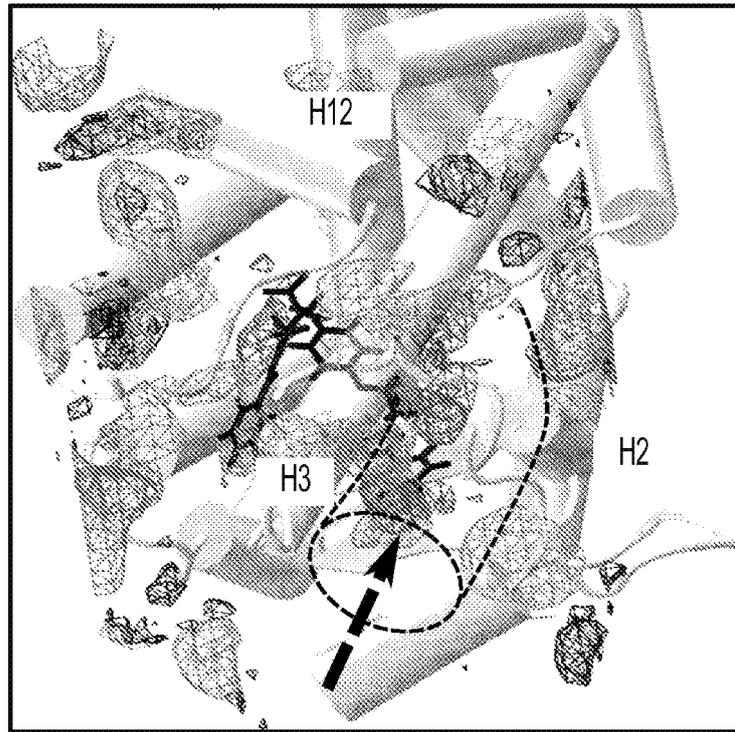
GW409544

**FIG. 14C**

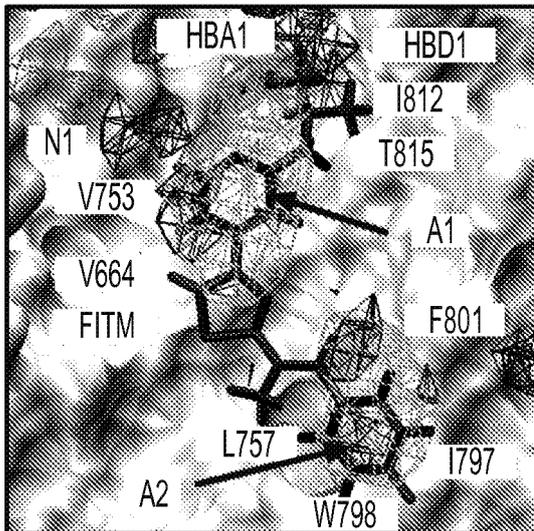


CERCO-A

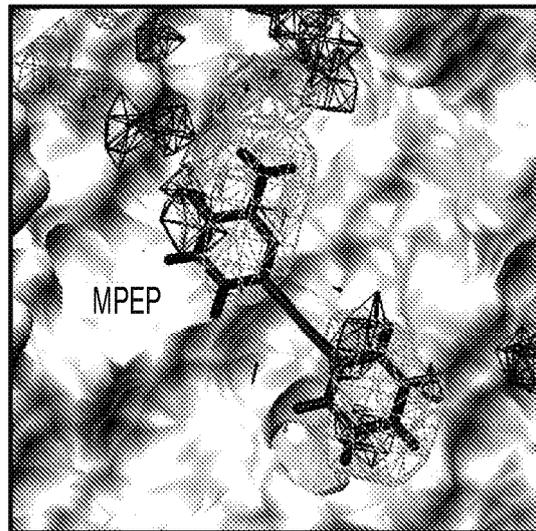
**FIG. 14D**



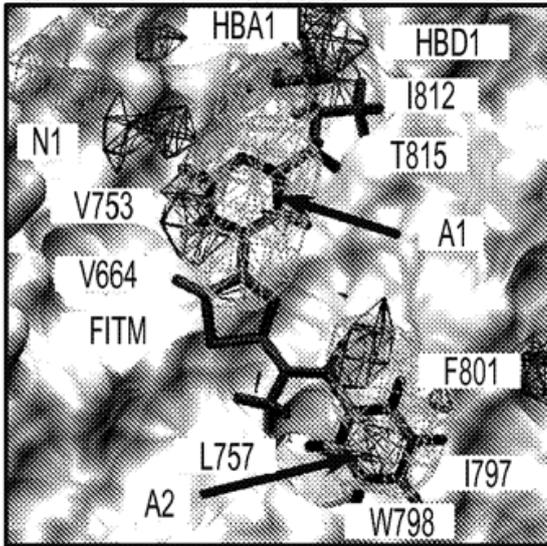
**FIG. 15**



**FIG. 16A**

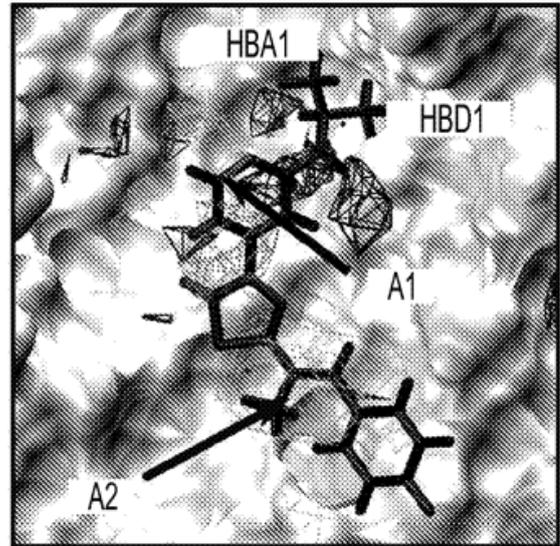


**FIG. 16B**



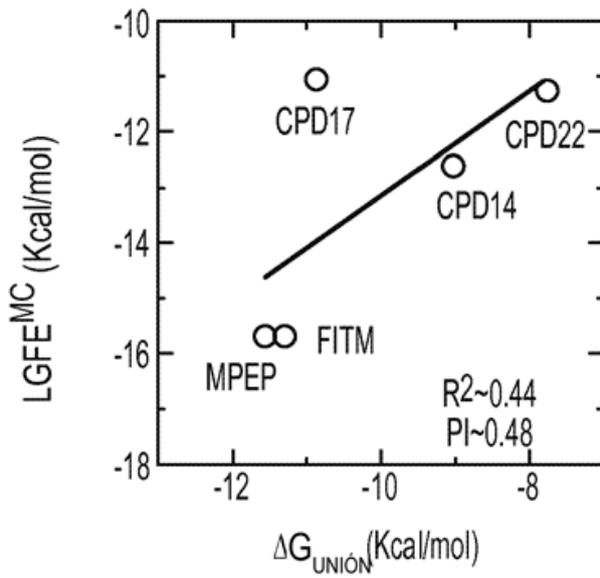
GCMC/MD

**FIG. 17A**

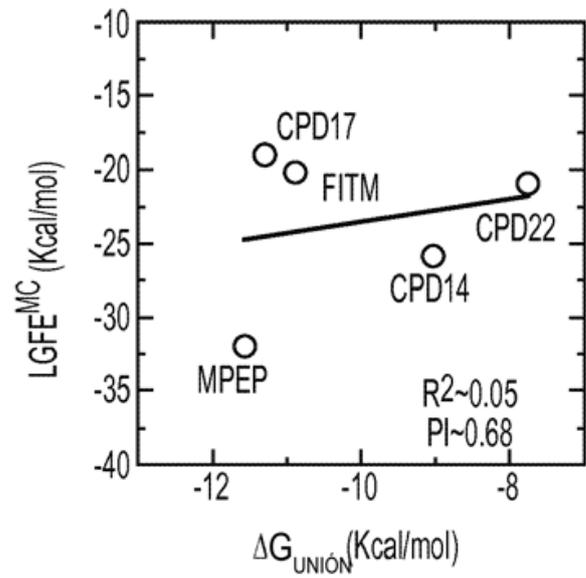


SÓLO DE GCMC

**FIG. 17B**

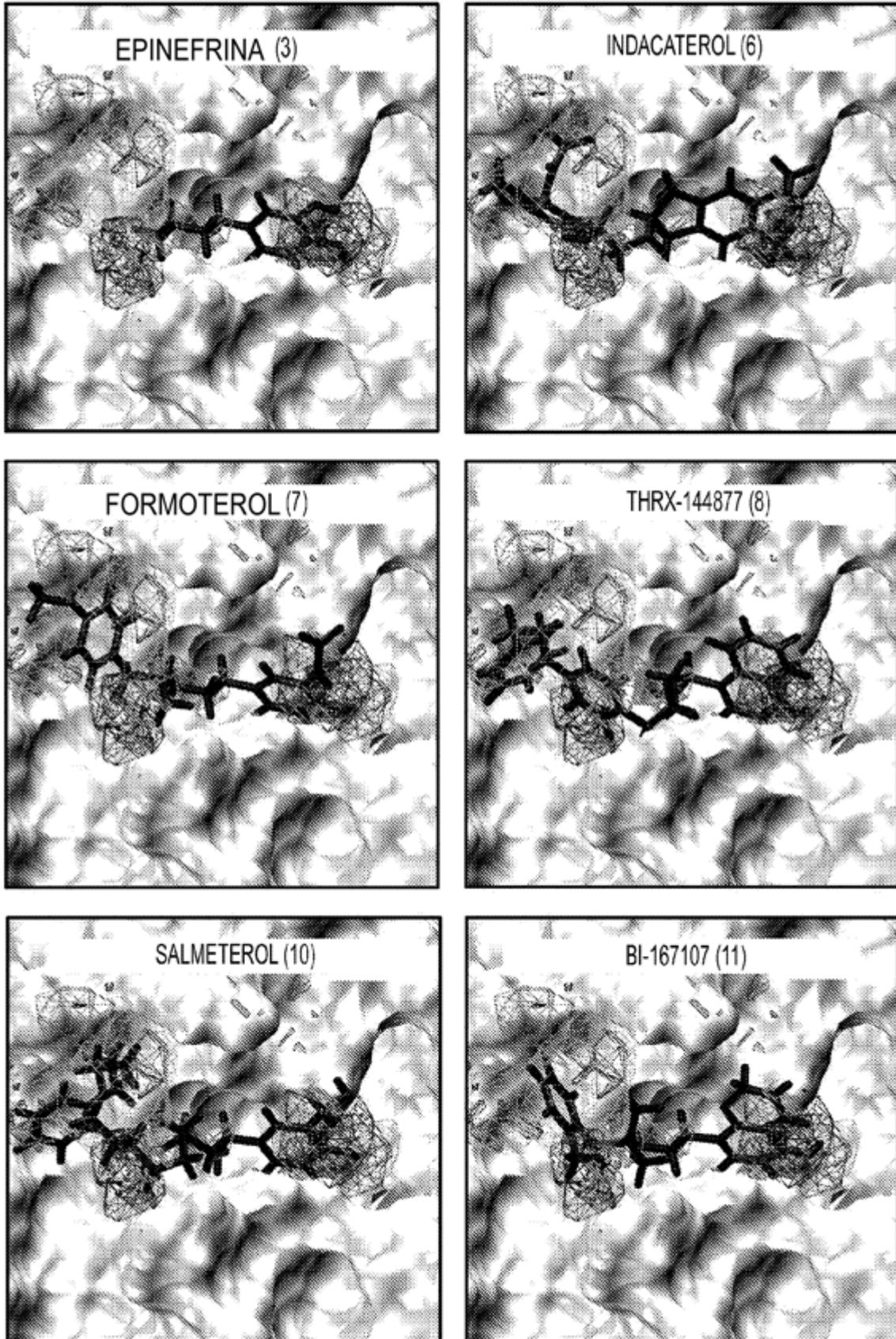


**FIG. 17C**

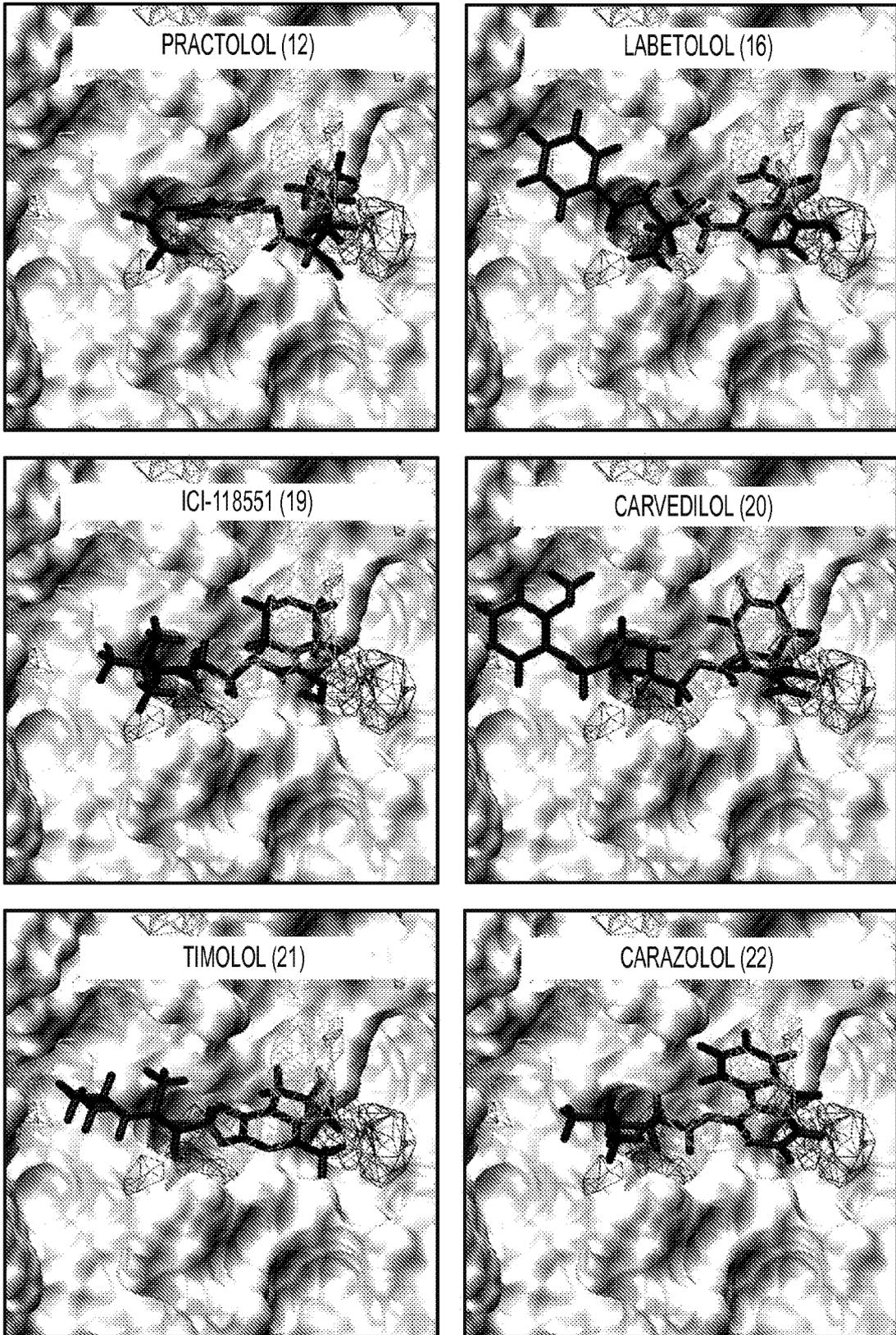


**FIG. 17D**

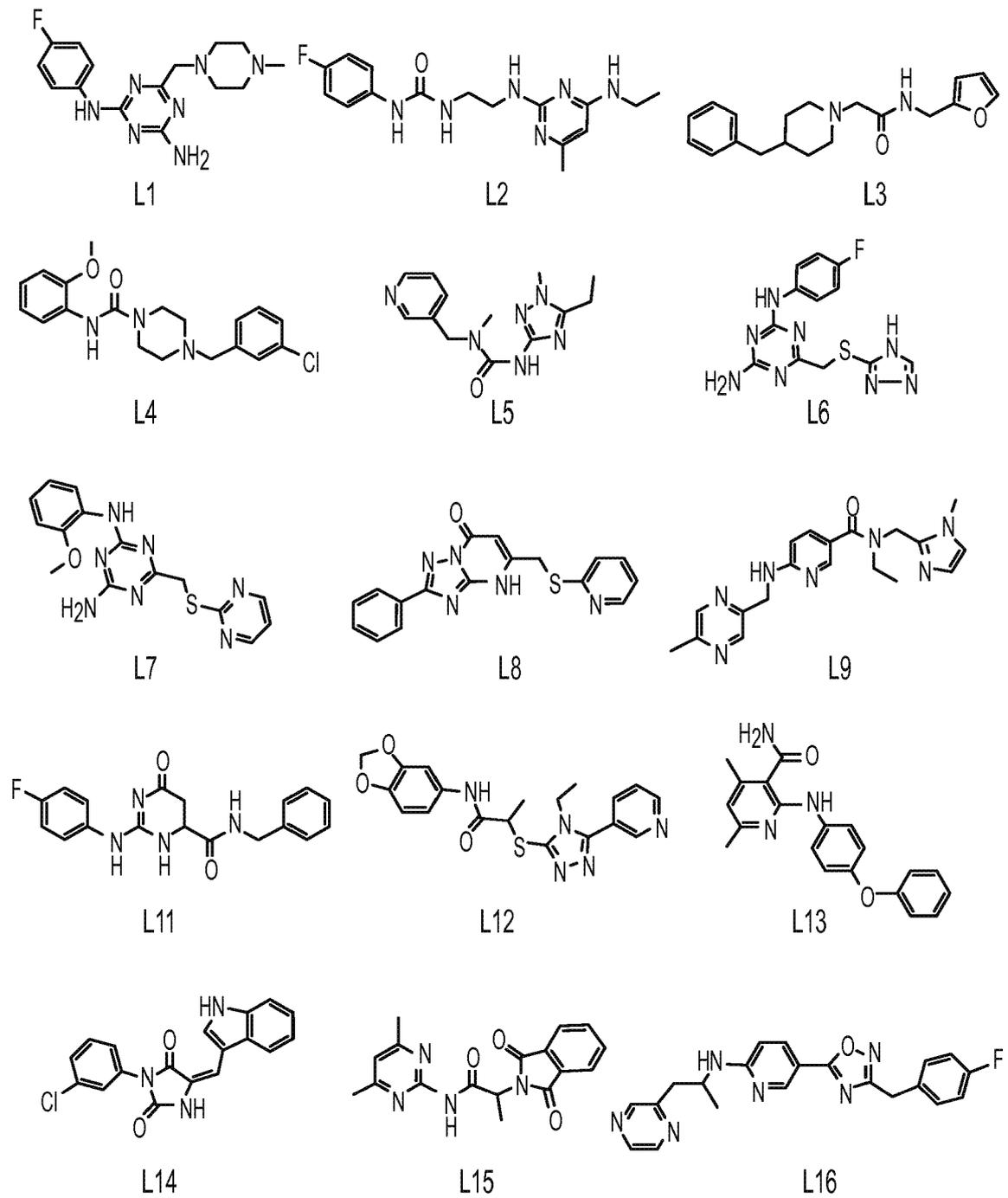




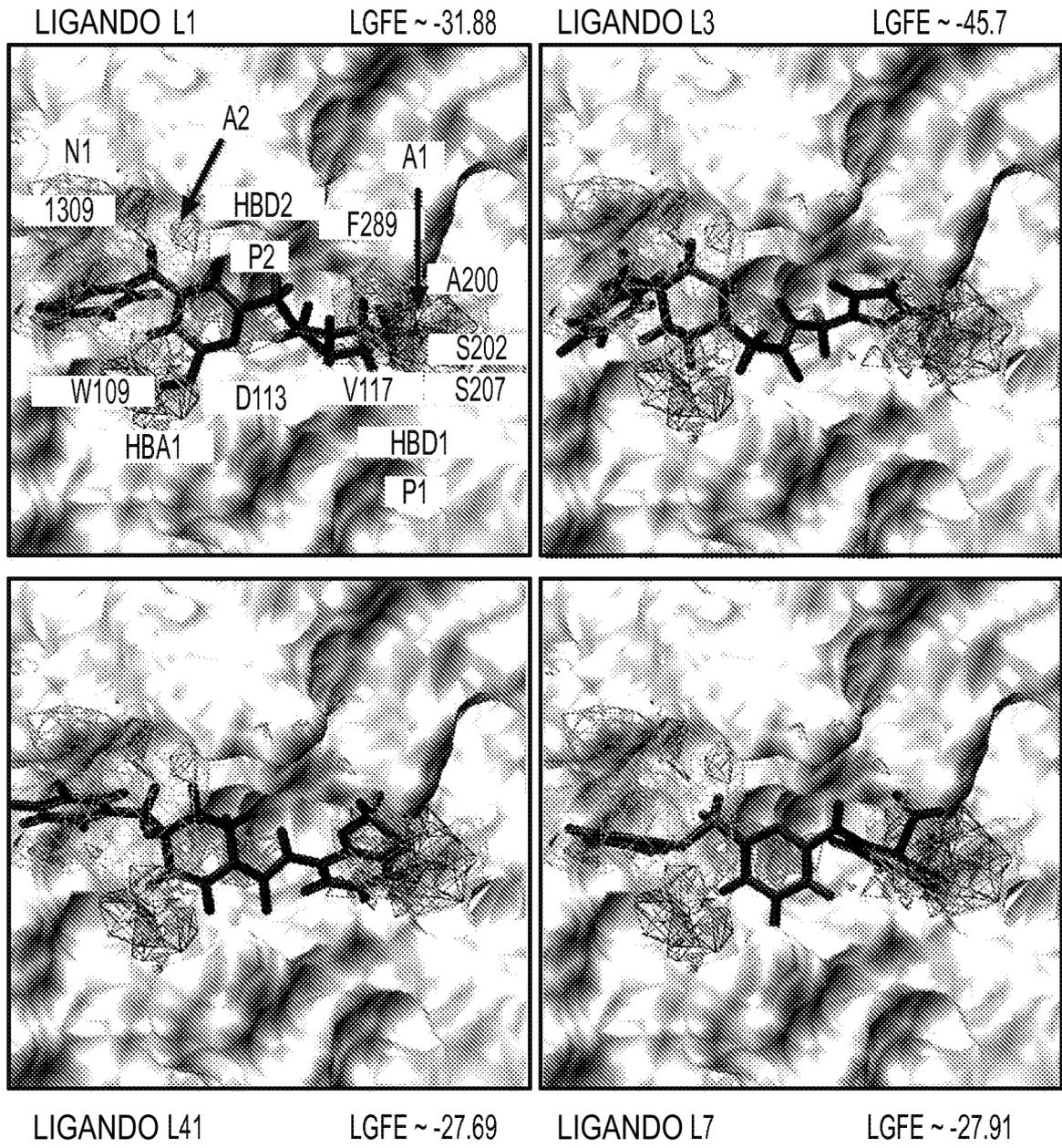
**FIG. 19**



**FIG. 20**



**FIG. 21**



**FIG. 22**