

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 853 348**

51 Int. Cl.:

G16B 15/30

(2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **13.03.2014** **PCT/US2014/025425**

87 Fecha y número de publicación internacional: **25.09.2014** **WO14151310**

96 Fecha de presentación y número de la solicitud europea: **13.03.2014** **E 14769374 (1)**

97 Fecha y número de publicación de la concesión europea: **11.11.2020** **EP 2972311**

54 Título: **Estimación de cierres de ciclos de afinidades de unión relativas y errores**

30 Prioridad:

15.03.2013 US 201313840039

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

15.09.2021

73 Titular/es:

SCHRODINGER, LLC (100.0%)
120 West Forty-Fifth Street, 17th Floor Tower 45
New York, New York 10036-4041, US

72 Inventor/es:

WANG, LINGLE;
LIN, TENG y
ABEL, ROBERT

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 853 348 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

Estimación de cierres de ciclos de afinidades de unión relativas y errores

REIVINDICACIÓN DE PRIORIDAD

La presente solicitud reivindica prioridad a la solicitud de patente de EE. UU. Nº de serie 13/840.039, presentada el 15 de marzo de 2013.

CAMPO TÉCNICO

La presente invención está en el campo general de la bioinformática, más específicamente la estimación o el cálculo informático de afinidades de unión relativas de receptor/ligando y sus errores.

ANTECEDENTES

Los procesos biológicos dependen frecuentemente de eventos de unión proteína-ligando (por ejemplo, la unión de un ligando a su receptor), y así el cálculo preciso de la energética asociada es un objetivo fundamental del diseño informático de fármacos basado en la estructura.¹⁻³

Se ha desarrollado una variedad de enfoques diferentes para calcular las energías libres de unión proteína-ligando con diferentes compensaciones entre la velocidad y la exactitud, que incluyen los métodos de punto final rápido tales como las funciones de puntuación empírica en el cribado virtual y la mecánica molecular/área superficial de nacimiento generalizado (MM/GBSA) o los modelos de Poisson-Boltzmann (MM/PBSA), el método de WM/MM que considera más explícitamente la contribución de la desolvatación de proteínas que los métodos de GB, y los métodos de perturbación de energía libre (FEP) e integración termodinámica (TI) diseñados para considerar una descripción energética completa o esencialmente completa de la unión dentro de los límites de exactitud del campo de fuerza subyacente y muestreo completo.^{1,2,4-10}

Entre los diversos métodos para calcular las afinidades de unión proteína-ligando, los cálculos de energía libre, tales como TI, FEP, dinámica lambda, OSRW alquímicos, etc., proporcionan generalmente una descripción termodinámicamente completa del evento de unión y, al menos en teoría, predicciones precisas del rendimiento dentro de los límites del campo de fuerza usado y muestreo completo del espacio de fase. Sin embargo, la aplicación práctica de los cálculos de energía libre en contextos industriales se puede limitar por el gran recurso informático requerido, particularmente cuando el muestreo de conformaciones múltiples, las escalas de tiempo necesarias para afectar los proyectos de vida y la necesidad de un enfoque robusto y significativo para determinar el error de muestreo asociado a dichos cálculos de energía libre.^{4, 11, 12}

El documento de patente US 2010/0241411 describe una técnica para enumerar exhaustivamente las propiedades termodinámicas de las moléculas de agua que solvatan el sitio activo de una proteína en su estado apo y calcular las afinidades de unión relativas de compuestos congéneres que se unen a esta proteína.

Wang et al. (PNAS 109:6) describen la aplicación de la perturbación de energía libre y la perturbación de energía libre/intercambio de réplicas con atemperado de solutos a dos modificaciones de complejos proteína-ligando que conducen a cambios conformacionales significativos.

Cozzini et al. (Curr. Med. Chem. 11:10) describen diversos métodos de evaluación de la contribución de las moléculas de agua a la energía libre de ligando que se une a proteínas.

Beveridge et al., Ann. Rev. Biophys. Bioeng. 18:1) revisan cálculos de energía libre por simulaciones moleculares.

SUMARIO

Los presentes inventores han descubierto un método generalizable implementado por ordenador para mejorar los cálculos de energía libre de afinidades de unión relativas de receptor-ligando, particularmente para estimar las afinidades de unión relativas de conjuntos congéneres de ligandos y determinar los errores de muestreo asociados a esas estimaciones.

Un aspecto de la invención caracteriza, en general, un método implementado por ordenador de determinación de la intensidad relativa de la unión entre un receptor y miembros individuales de un conjunto de ligandos para formar complejos entre miembros individuales del conjunto de ligandos y el receptor. El método incluye una determinación implementada por ordenador de múltiples diferencias de energía libre de unión relativas para un conjunto de pares de ligandos que forman al menos un ciclo cerrado, y una determinación implementada por ordenador de la magnitud de la histéresis alrededor de cada ciclo termodinámico cerrado. En general, el método incluye,

a. una determinación probabilística implementada por ordenador de las diferencias de energía libre y distribuciones de error alrededor de las diferencias de energía libre a lo largo de cada una de las ramas del ciclo termodinámico cerrado que conducen probabilísticamente al valor de histéresis observado para los ciclos termodinámicos cerrados;

b. una determinación implementada por ordenador de la diferencia de energía libre más probable para cada rama en el ciclo termodinámico cerrado incluida en el modelo probabilístico determinado en a;

c. una determinación implementada por ordenador del error más probable asociado a la diferencia de energía libre de unión más probable para cada par de ligandos a lo largo de cada rama en el ciclo termodinámico cerrado de la determinación probabilística en b.;

d. una salida implementada por ordenador de valores que representan las energías libres de unión relativas y errores en c. a un usuario del ordenador.

En realizaciones preferidas, la etapa de estimar el error comprende el análisis implementado por ordenador de diferencias de energía libre de unión entre ligandos a lo largo de las ramas de más de un ciclo termodinámico cerrado, y la determinación implementada por ordenador de la magnitud de la histéresis alrededor de cada uno de los ciclos termodinámicos cerrados. También en realizaciones preferidas, las etapas c y d incluyen determinar un conjunto de valores de energía libre para cada rama que minimiza la función

$$\sum_i \frac{(E_i - F_i)^2}{2\sigma_i^2}$$

con la limitación:

$$\sum_i F_i = 0$$

donde E_i es diferencia de energía libre calculada para una rama i dada;

F_i es la diferencia de energía libre teórica para una rama i dada;

y σ_i es la desviación estándar de las diferencias de energía libre calculadas para la rama i , y donde la suma de las diferencias de energía libre teóricas para todos los ciclos cerrados es 0.

En otras realizaciones preferidas: el receptor es una proteína; los ligandos son congéneres; se supone una distribución gaussiana en la construcción del modelo probabilístico de la histéresis observada en la etapa a.; se supone que la distribución del error asociada a las simulaciones de energía libre en la etapa a. es uniforme; se supone que la distribución de error se suma al error de Bennett en la etapa a. La conectividad de los ciclos termodinámicos cerrados se representa como un gráfico o como una matriz. La determinación probabilística se puede hacer por diversos métodos, que incluyen, sin limitación, métodos teóricos gráficos, métodos algebraicos matriciales, métodos bayesianos, métodos de máxima verosimilitud.

Otros aspectos de la invención caracterizan un medio legible por ordenador que comprende instrucciones no transitorias tangibles para realizar los métodos anteriormente descritos, y una unidad de procesamiento gráfica de uso general con instrucciones no transitorias legibles por ordenador para realizar los métodos.

Más adelante, los presentes inventores ejemplifican la invención de la estimación de la afinidad del ciclo de cierre y la maquinaria de análisis de error para FEP/REST (perturbación de energía libre/intercambio de réplicas con atemperado de solutos),¹²⁻¹⁵ para FEP/MD y para cálculos de energía libre de afinidad de unión relativa de TI, pero la invención se puede usar con varias otras plataformas de cálculo de energía libre de unión relativa, que incluyen dinámica lambda, metadinámica química, OSRW químico u otros. La invención también se puede emplear independientemente de si se implementa la plataforma de cálculo de energía libre relativa particular empleando algoritmos de muestreo basados en dinámica molecular, monte carlo, monte carlo gran canónico, dinámica molecular con intercambio de réplicas, dinámica molecular acelerada, o cualquier otro protocolo de muestreo.

En particular, los presentes inventores ejemplifican la invención investigando un subconjunto de una serie de ligandos congéneres que se unen al receptor de cinasa dependiente de ciclina CDK2-ciclina A.¹⁶ CDK2 es un miembro de la familia de CDK que realiza diversas funciones en la regulación de la proliferación celular y los ciclos de transcripción de ARN polimerasa II. CDK2 también se ha identificado como una diana importante de fármacos para las estrategias terapéuticas selectivas de tumor.^{17,18}

A pesar de los sistemas específicos que presentan los presentes inventores para ejemplificar la invención, los expertos en la técnica entenderán que la invención se puede usar como una herramienta general para la investigación de unión a ligando. Para ilustrar este punto, los presentes inventores también han aplicado la invención al cálculo de las afinidades de unión relativas de ligandos que se unen a JNK cinasa y BACE, que también son dianas de proteína de interés médico significativo.

Los detalles de una o más realizaciones de la invención se exponen en los dibujos adjuntos y en la descripción que sigue. Otras características, objetos y ventajas de la invención serán evidentes a partir de la descripción y los dibujos, y de las reivindicaciones.

DESCRIPCIÓN DE LOS DIBUJOS

El archivo de patente o de solicitud contiene al menos un dibujo realizado en color. Las copias de esta publicación de patente o de solicitud con dibujo(s) a color se proporcionarán por la oficina previa solicitud y pago de la tasa necesaria.

5 **Figura 1:** Un modelo simple que ilustra el problema de la comprobación de la coherencia y la fiabilidad en los cálculos de FEP. Para un conjunto de tres ligandos L1, L2 y L3, se realizaron tres cálculos de FEP con diferencias de energía libre estimadas E21, E32 y E13, respectivamente. ¿Cómo lograr las mejores estimaciones de las energías libres relativas (F21, F32 y F13) a partir de los tres cálculos de FEP, y cómo determinar si los cálculos son fiables?

Figura 2: Una ruta más complicada con dos ciclos que comparten la misma arista usada en el cálculo de la afinidad de unión de los inhibidores de CDK2.

10 **Figura 3:** Ejemplo gráfico de los cálculos de energía libre relativa en aplicaciones generales. Cada vértice en el gráfico representa un ligando. Cada arista en el gráfico representa cálculos de energía libre relativa, la flecha de la arista indica la dirección de cálculo de la energía libre de unión relativa y el valor en cada arista es la diferencia de energía libre calculada del cálculo de energía libre para ese par particular de ligandos.

15 **Figura 4:** Descomposición del gráfico completo en la Figura 3 en 3 subgráficos independientes. Los subgráficos pueden compartir el mismo vértice, pero no compartirán la misma arista.

20 **Figura 5:** Las estructuras de 6 ligandos de CDK2 con estructuras cristalinas y las rutas usadas para los cálculos de FEP/REST y FEP/MD. Las tres flechas rojas representan las rutas de cálculo de la energía libre de unión relativa que completan tres ciclos distintos para evaluar la coherencia y la fiabilidad de los cálculos. Se usó el ligando 1h1q para los 5 cálculos de energía libre de unión relativa azules puesto que los otros compuestos se pueden generar todos simplemente añadiendo diferentes sustituyentes en el anillo de benceno del ligando 1h1q.

25 **Figura 6:** Los resultados predichos por FEP/MD y FEP/REST frente a los datos de afinidad de unión experimentales para el conjunto de inhibidores de CDK2 con estructuras cristalinas. Las afinidades de unión predichas por FEP/MD no varían significativamente, que hace que sea imposible ordenar de forma fiable los ligandos 1h1r, 1h1s, 1oiu, 1oi9 y 1oiy usando estos valores. Por el contrario, los resultados predichos por FEP/REST tienen una correlación muy alta con los datos experimentales ($R^2=0,91$), y la desviación es inferior a 0,8 kcal/mol para todos los pares. Las barras de error para las predicciones de FEP/REST informadas en esta figura se calculan a modo de las realizaciones de la invención usando análisis de cierre de ciclos.

30 **Figura 7:** La conformación inicial correspondiente a la estructura cristalina para la simulación de FEP/REST para el ligando 1h1r (izquierda), el modo de unión alternativo, con el anillo de fenilo sustituido con cloro dado la vuelta, observado en la ruta de REST/REST (centro) y la densidad electrónica del ligando 1h1r en la cristalografía de rayos X (derecha). Las proteínas se representan por cintas coloreadas para el esqueleto y la representación de varillas para las cadenas laterales, mientras que los ligandos se representan en representación en tubo, y la densidad electrónica se presenta como una superficie amarilla sólida. El corte sigma usado para la visualización de la densidad es 1,0. Para el ligando 1h1r, la densidad electrónica tiene dos picos para el anillo de benceno sustituido con cloro, que indica que el ligando tiene dos modos de unión, de acuerdo con las simulaciones de FEP/REST.

35 **Figura 8:** La conformación inicial correspondiente a la estructura cristalina para la simulación de FEP/REST para el ligando 1h1s (izquierda) y el modo de unión alternativo observado en la ruta de REST/REST (derecha). Las proteínas se representan por cintas coloreadas para el esqueleto y la representación de varillas para las cadenas laterales, mientras que los ligandos se representan en representación en tubo. En comparación con la conformación inicial, el grupo sulfonamida gira aproximadamente 120° alrededor del enlace C-S, y los restos que rodean el sitio de unión, especialmente Gln85 y Asp86, se reordenan en consecuencia. Se forman dos enlaces de hidrógeno entre el grupo sulfonamida y los restos de proteína.

Figura 9: Las rutas de cálculo de la energía libre de unión relativa usadas para los cálculos de FEP/REST para los ligando sin estructura cristalina que se unen al receptor de CDK2.

45 **Figura 10:** Los resultados predichos de FEP/REST frente a los datos de afinidad de unión experimentales para el conjunto completo de inhibidores de CDK2 que incluyen tanto los 6 ligandos con estructuras cristalinas como los 10 ligandos sin estructuras cristalinas. Los 6 ligandos con estructuras cristalinas (presentados en la Figura 6) están coloreados de rojo en la figura y los 10 ligandos sin estructuras cristalinas están coloreados de negro. Las barras de error para las predicciones de FEP/REST informadas en esta figura se calculan basándose en el método detallado en el Apéndice A usando cierres de ciclo.

Figura 11: Rutas de cálculo de la energía libre de unión relativa para los 7 ligandos que se unen a los receptores de Jnk1 con 2 cierres de ciclo independientes.

Figura 12: Rutas de cálculo de la energía libre de unión relativa para los 6 ligandos que se unen a BACE con dos cierres de ciclo independientes.

Figura 13: Rutas de cálculo de la energía libre de unión relativa para los 3 que se unen a escitalona deshidratasa con y sin agua unida en el sitio activo. Las simulaciones con agua unida en el sitio activo se indican por '(H2O)'. Se construyen 4 cierres de ciclo independientes en las rutas de cálculo de la energía libre de unión relativa.

Figura 14: Ciclos termodinámicos que conectan afinidades de unión metano-placa y las energías libres de cargar las placas en agua. Las partículas grises representan los átomos de Lennard Jones que forman el recinto hidrófobo modelo, las partículas rojas representan iones negativamente cargados, las partículas azules representan iones positivamente cargados y las partículas verdes representan metano del átomo unido, que se une a los recintos.

Figura 15: Un diagrama de bloques de un sistema informático que realiza el método.

Figura 16: Un diagrama de flujo de las etapas en el método.

10 DESCRIPCIÓN DETALLADA

Las afinidades de unión relativas calculadas con la invención usando cierre de ciclo se correlacionan mucho con los datos experimentales, y se produce una barra de error significativa que indica la convergencia de los cálculos.

Esto es especialmente evidente cuando se emplea la invención para evaluar los cálculos de energía libre de FEP/REST, que proporciona un muestreo exhaustivo de las diferentes conformaciones del ligando en el sitio activo del receptor. Aunque no se desea quedar ligado a teoría particular, los presentes inventores identificaron que para dos de los ligandos existen múltiples modos de unión y que el muestreo de aquellos modos es importante para la correcta predicción de las afinidades de unión. Aunque FEP/REST puede muestrear las importantes conformaciones en un tiempo de simulación relativamente corto, los ligandos se atraparon en las conformaciones iniciales usando FEP/MD. El método ilustra esta característica, que indica un alto error asociado a los cálculos de energía libre de FEP/MD, y un bajo error con los cálculos de FEP/REST. Así, la invención puede mejorar las técnicas de cálculo de energía libre relativa proporcionando una estimación razonable de los errores asociados en el cálculo y la fiabilidad de las predicciones.¹⁹

Inicialmente, el método construye cierres de ciclo en cálculos de energía libre relativa como se ha hecho por varios otros grupos,^{10,20,21} entonces desde esta entrada, el método determina la coherencia y la fiabilidad de los cálculos de energía libre relativa evaluando cuándo se desvía la suma de los cambios de energía libre calculada para cada ciclo termodinámico cerrado del valor teórico de 0. En la sección que sigue, los presentes inventores describen el método con detalle, de forma que un experto en la técnica puede ver inmediatamente cómo el método da predicciones *fiabiles* de las energías libres relativas de las múltiples estimaciones de energía libre obtenidas a modo de los cálculos de energía libre relativa del cierre de ciclo, así como los límites de error esperados asociados a los cálculos, así como el mecanismo para marcar cálculos de energía libre relativa que tienen errores sistemáticos.

Especificación detallada de la invención de coherencia de energía libre relativa de cierre de ciclo

1. Validación de coherencia y el mejor estimador de energía libre usando cierre de ciclo en rutas de cálculo de energía libre de unión relativa

El siguiente modelo simple ilustra los métodos y los conceptos del método. Considérese un conjunto de tres ligandos, L_1 , L_2 y L_3 . Supóngase que las diferencias de energía libre de unión relativa experimentalmente medidas entre los tres ligandos son $F_{21}^{\text{exp}} = F_2^{\text{exp}} - F_1^{\text{exp}}$, $F_{32}^{\text{exp}} = F_3^{\text{exp}} - F_2^{\text{exp}}$ y $F_{13}^{\text{exp}} = F_1^{\text{exp}} - F_3^{\text{exp}}$, donde F_1^{exp} , F_2^{exp} y F_3^{exp} son las energías libres de unión experimentalmente medidas para los ligandos L_1 , L_2 y L_3 , respectivamente. La energía libre es una propiedad termodinámica, por tanto $F_{13}^{\text{exp}} + F_{32}^{\text{exp}} + F_{21}^{\text{exp}} = 0$ (véase la Figura 1). Se supone que se hacen tres cálculos de energía libre de unión relativa, de L_1 a L_2 , L_2 a L_3 y L_3 a L_1 , y las diferencias de energía libre calculadas para las tres rutas de cálculo de energía libre de unión relativa son E_{21} , E_{32} y E_{13} , respectivamente. Si las simulaciones convergen perfectamente y el campo de fuerza es perfecto, entonces idealmente $E_{21} = F_{21}^{\text{exp}}$, $E_{32} = F_{32}^{\text{exp}}$, $E_{13} = F_{13}^{\text{exp}}$, y $E_{13} + E_{21} + E_{32} = 0$. En la práctica, sin embargo, existen errores asociados a las energías libres relativas calculadas y, en general, $E_{21} \neq F_{21}^{\text{exp}}$, $E_{32} \neq F_{32}^{\text{exp}}$, $E_{13} \neq F_{13}^{\text{exp}}$ y $E_{13} + E_{21} + E_{32} = \Delta \neq 0$. Los presentes inventores lo llaman Δ , la histéresis del cálculo de energía libre asociada a este cierre de ciclo. Estos errores incluyen los errores estadísticos no sesgados debido a las fluctuaciones al azar y los errores sesgados debido al muestreo incompleto del espacio de fase (la proteína y/o el ligando están atrapados en un mínimo local en el espacio de conformación) y el error en el campo de fuerza. El método evalúa la coherencia y la fiabilidad de estos cálculos.

Los errores en los cálculos de energía libre relativa en comparación con los valores experimentales se pueden separar en dos categorías: el error sistemático procedente de la diferencia del campo de fuerza usado en la simulación en comparación con la verdadera superficie de energía potencial del sistema, y el error procedente de la no convergencia, ya sea debido a muestreo aleatoriamente o sistemáticamente incompleto del espacio de fase, o del estimador de energía libre en sí, por ejemplo, TI frente a FEP, etc. F_{21} indica la diferencia de energía libre teórica entre dos estados

termodinámicos (es decir, de ligando L_1 a ligando L_1) para el campo de fuerza subyacente de una simulación no sesgada larga infinita y estimador de energía libre relativa no sesgada. Si no existe error sistemático en el campo de fuerza, entonces $F_{21} = F_{21}^{\text{exp}}$. En los cálculos de energía libre relativa prácticos, por ejemplo para el cálculo de L_1 a L_2 , la simulación se realiza con una cantidad finita de tiempo de muestreo y el muestreo puede tener cierto sesgo, así la energía libre calculada E_{21} se podría desviar de su valor teórico F_{21} , de forma que dependa de la configuración inicial de la simulación.

Repitiendo el mismo cálculo de energía libre relativa un número infinito de veces a partir de diferentes configuraciones iniciales y diferentes semillas aleatorias para las velocidades, las energías libres calculadas tienen una distribución. Sin pérdida de generalidad, para el cálculo de la energía libre de unión relativa de ligando L_i a ligando L_j con diferencia de energía libre teórica entre ellos F_{ji} , $P(E_{ji} | F_{ji})$ indica la distribución de la energía libre calculada E_{ji} . $P(E_{ji} | F_{ji})$ puede ser en principio cualquier tipo de distribución, tales como la distribución gaussiana, la distribución de Lorentz, la distribución uniforme, la distribución delta, etc.

$$P(E_{ji} | F_{ji}) = \begin{cases} \delta(E_{ji} - F_{ji}) & \text{distribución } \delta \\ \frac{1}{2a} & \text{cuando } E_{ji} \in [F_{ji} - a, F_{ji} + a] \text{ distribución uniforme} \\ \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(E_{ji} - F_{ji})^2}{2\sigma^2}\right) & \text{distribución gaussiana} \\ \dots & \text{otras distribuciones} \end{cases}$$

Este método predice la diferencia de energía libre del ligando L_i al ligando L_j y el error asociado para la predicción basada en la distribución $P(E_{ji} | F_{ji})$ con cierres de ciclo contruados en las rutas de cálculo de energía libre de unión relativa, es decir:

$$P(E_{ji} | F_{ji})|_{\text{ciclos cerrados en las rutas}} \rightarrow \{\hat{E}_{ji}, e_{ji}\}$$

donde $\hat{E}_{ji}$ indica la diferencia de energía libre predicha entre el ligando L_i y el ligando L_j , y e_{ji} indica el error asociado a la predicción.

La predicción se puede hacer usando muchos métodos diferentes, por ejemplo, el método de máxima verosimilitud, que maximiza la probabilidad de la observación; el método estadístico bayesiano, que optimiza los parámetros basados en las energías libres calculadas, etc.

Aquí, los presentes inventores ejemplifican la inversión mediante un ejemplo específico, que usa el método de máxima verosimilitud y que supone distribución gaussiana para las energías libres calculadas. Los expertos en la técnica entenderán que la invención se puede usar como una herramienta general para otros tipos de métodos de análisis estadístico y otros tipos de distribuciones. Al final de esta sección también se da una derivación de ejemplo usando estadística bayesiana.

Se supone que las energías libres calculadas están distribuidas de forma gaussiana con F_{21} promedio (sin sesgo sistemático) y desviación estándar σ_{21} . Entonces la densidad de probabilidad de que un único cálculo de energía libre relativa proporcione un valor de E_{21} para esta ruta es:

$$\rho_{21}(E_{21} | F_{21}) = \frac{1}{\sqrt{2\pi}\sigma_{21}} \exp\left(-\frac{(E_{21} - F_{21})^2}{2\sigma_{21}^2}\right)$$

Similarmente, para las rutas L_2 a L_3 , y L_3 a L_1 , la densidad de probabilidad de que los cálculos de energía libre relativa proporcionen valores de E_{32} y E_{13} es respectivamente:

$$\rho_{32}(E_{32} | F_{32}) = \frac{1}{\sqrt{2\pi}\sigma_{32}} \exp\left(-\frac{(E_{32} - F_{32})^2}{2\sigma_{32}^2}\right)$$

$$\rho_{13}(E_{13} | F_{13}) = \frac{1}{\sqrt{2\pi}\sigma_{13}} \exp\left(-\frac{(E_{13} - F_{13})^2}{2\sigma_{13}^2}\right)$$

Para un conjunto dado de diferencias de energía libre teóricas F_{21} , F_{32} y F_{13} , la verosimilitud global, L , de que los tres cálculos de energía libre relativa proporcionen valores de E_{21} , E_{32} y E_{13} es:

$$L(E_{21}, E_{32}, E_{13} | F_{21}, F_{32}, F_{13}) = \frac{1}{(2\pi)^{3/2} \sigma_{21} \sigma_{32} \sigma_{13}} \exp\left(-\sum_i \frac{(E_i - F_i)^2}{2\sigma_i^2}\right) \text{ donde } i \in (21, 32, 13)$$

Según el método de máxima verosimilitud, los valores más probables de F_{21} , F_{32} y F_{13} son el conjunto de valores que maximizan la verosimilitud anterior. Haciendo el logaritmo de la expresión anterior, el conjunto de valores que maximiza la verosimilitud es el conjunto de valores que minimizan la siguiente función:

$$\sum_i \frac{(E_i - F_i)^2}{2\sigma_i^2} \text{ donde } i \in (21, 32, 13)$$

con la limitación:

$$F_{21} + F_{32} + F_{13} = 0$$

Usando multiplicadores de Lagrange, el conjunto de valores que maximiza la verosimilitud es:

$$\begin{aligned} \hat{F}_{21} &= E_{21} - \frac{\sigma_{21}^2}{\sigma_{21}^2 + \sigma_{32}^2 + \sigma_{13}^2} (E_{21} + E_{32} + E_{13}) = \frac{\sigma_{32}^2 + \sigma_{13}^2}{\sigma_{21}^2 + \sigma_{32}^2 + \sigma_{13}^2} E_{21} - \frac{\sigma_{21}^2}{\sigma_{21}^2 + \sigma_{32}^2 + \sigma_{13}^2} (E_{32} + E_{13}) \\ \hat{F}_{32} &= E_{32} - \frac{\sigma_{32}^2}{\sigma_{21}^2 + \sigma_{32}^2 + \sigma_{13}^2} (E_{21} + E_{32} + E_{13}) = \frac{\sigma_{21}^2 + \sigma_{13}^2}{\sigma_{21}^2 + \sigma_{32}^2 + \sigma_{13}^2} E_{32} - \frac{\sigma_{21}^2}{\sigma_{21}^2 + \sigma_{32}^2 + \sigma_{13}^2} (E_{21} + E_{13}) \\ \hat{F}_{13} &= E_{13} - \frac{\sigma_{13}^2}{\sigma_{21}^2 + \sigma_{32}^2 + \sigma_{13}^2} (E_{21} + E_{32} + E_{13}) = \frac{\sigma_{32}^2 + \sigma_{21}^2}{\sigma_{21}^2 + \sigma_{32}^2 + \sigma_{13}^2} E_{13} - \frac{\sigma_{13}^2}{\sigma_{21}^2 + \sigma_{32}^2 + \sigma_{13}^2} (E_{32} + E_{21}) \end{aligned}$$

Los estimadores anteriores no tienen sesgo sistemático, y no habrá discrepancia sustancial entre las predicciones de energía libre de las diferentes rutas. El método interpreta los estimadores anteriores como el promedio ponderado de dos rutas. Por ejemplo, se puede estimar la diferencia de energía libre entre el ligando L_1 y L_2 , F_{21} , de E_{21} , o de $-(E_{32} + E_{13})$, y el mejor estimador es un promedio ponderado de las dos predicciones. Cuanto más pequeña sea la desviación estándar para la diferencia de energía libre calculada a lo largo de la ruta, más grande es el peso para el mejor estimador, y viceversa.

Además, según el modelo anterior, la histéresis del cierre de ciclo, $E_{21} + E_{32} + E_{13} = \Delta$, también se distribuye de forma gaussiana con promedio de 0 y desviación estándar de $s = \sqrt{\sigma_{21}^2 + \sigma_{32}^2 + \sigma_{13}^2}$. Si la suma de las tres energías libres calculadas se desvía más de 2s desde 0, es casi cierto ($P = 0,95$) que los cálculos no convergen y los resultados no son fiables; si la suma de las tres energías libres calculadas se desvía más de s desde 0, es muy probable ($P = 0,68$) que el cálculo no converja y los resultados no pueden ser fiables. Así, el método determina si las predicciones son fiables.

En los cálculos de energía libre relativa prácticos, debido al alto coste informático, normalmente solo se realiza un cálculo dado una vez, y la desviación estándar no se puede estimar para cada predicción. En este caso, el método supone que la desviación estándar para cada cálculo es la misma, por ejemplo 0,8 kcal/mol para cada ruta. Entonces, si la discrepancia de los resultados de la energía libre de dos rutas es mayor que 1,4 kcal/mol, indica que el cálculo no converge y probablemente no es fiable. En este supuesto, el mejor estimador de energía libre tiene la siguiente expresión:

$$\begin{aligned} \hat{F}_{21} &= E_{21} - \frac{1}{3} (E_{21} + E_{32} + E_{13}) = \frac{2}{3} E_{21} - \frac{1}{3} (E_{32} + E_{13}) \\ \hat{F}_{32} &= E_{32} - \frac{1}{3} (E_{21} + E_{32} + E_{13}) = \frac{2}{3} E_{32} - \frac{1}{3} (E_{21} + E_{13}) \\ \hat{F}_{13} &= E_{13} - \frac{1}{3} (E_{21} + E_{32} + E_{13}) = \frac{2}{3} E_{13} - \frac{1}{3} (E_{32} + E_{21}) \end{aligned}$$

Además, en el supuesto de que la desviación estándar para cada cálculo sea la misma, la histéresis del cierre de ciclo en sí proporciona una estimación del error asociado a cada predicción de energía libre. Como se trata anteriormente, la histéresis Δ del cierre de ciclo se distribuye de forma gaussiana con promedio de 0 y desviación estándar de $s = \sqrt{3}\sigma$. Así, la probabilidad de que un cálculo dado genere una histéresis con valor Δ para el cierre de ciclo es:

$$P = \frac{1}{\sqrt{2\pi}\sqrt{3}\sigma^2} \exp\left(-\frac{\Delta^2}{6\sigma^2}\right)$$

El valor de sigma que maximiza la probabilidad anterior da la máxima estimación de probabilidad de la desviación estándar asociada a cada cálculo de energía libre:

$$\hat{\sigma} = \frac{\Delta}{\sqrt{3}}$$

- 5 Obsérvese que en la derivación anterior, las energías libres y los errores asociados que el método estima son las estimaciones óptimas que se basan en toda la información obtenida de los cálculos de energía libre relativa. Si el campo de fuerza subyacente en los cálculos tiene sesgo sistemático, los resultados del cálculo de energía libre relativa perfectamente convergente se pueden desviar aún de los valores medidos experimentales, que es algo que no se puede corregir basándose en el análisis anterior.
- 10 El método que describen anteriormente los presentes inventores se aplica a casos más complicados con más miembros en el cierre de ciclo y varios ciclos que comparten la misma arista. Aquí, los presentes inventores ilustran el método de obtención de los mejores estimadores de energía libre para las rutas usadas en los cálculos de FEP/MD y FEP/REST para los 6 ligandos de CDK2 con estructuras cristalinas (Figura 2). El algoritmo más general para cualquier tipo de ruta de cálculo de la energía libre de unión relativa se da en la siguiente sección.

$$\begin{aligned} \hat{F}_{21} &= \frac{1}{2}E_{21} - \frac{1}{4}(E_{42} + E_{14}) - \frac{1}{4}(E_{32} + E_{13}) \\ \hat{F}_{32} &= \frac{5}{8}E_{32} - \frac{3}{8}(E_{13} + \frac{2}{3}E_{21} - \frac{1}{3}(E_{14} + E_{42})) \\ \hat{F}_{42} &= \frac{5}{8}E_{42} - \frac{3}{8}(E_{14} + \frac{2}{3}E_{21} - \frac{1}{3}(E_{13} + E_{32})) \\ \hat{F}_{13} &= \frac{5}{8}E_{13} - \frac{3}{8}(E_{32} + \frac{2}{3}E_{21} - \frac{1}{3}(E_{14} + E_{42})) \\ \hat{F}_{14} &= \frac{5}{8}E_{14} - \frac{3}{8}(E_{42} + \frac{2}{3}E_{21} - \frac{1}{3}(E_{13} + E_{32})) \\ \hat{F}_{43} &= \frac{1}{2}(E_{13} - E_{14}) + \frac{1}{2}(E_{42} - E_{32}) \end{aligned}$$

Mejores estimadores de energía libre usando estadística bayesiana

Para un cálculo de FEP que muta de ligando L_i a ligando L_j con la diferencia de energía libre teórica F_{ji} , supóngase que la energía libre calculada E_{ji} se distribuye de forma promedio con F_{ji} promedio y desviación estándar σ_{ji} (el mismo modelo que se usa para el método de máxima verosimilitud). Entonces, la densidad de probabilidad de que la energía libre calculada a partir de un cálculo de FEP proporcione un valor de E_{ji} es

$$\rho(E_{ji} | F_{ji}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(E_{ji} - F_{ji})^2}{2\sigma^2}\right)$$

Muchos valores diferentes de F_{ji} pueden conducir a la misma energía libre calculada E_{ji} con diferente posibilidad. Según la estadística bayesiana, antes de realizar el cálculo de FEP, el método tiene una estimación de la distribución (la distribución previa) de la diferencia de energía libre entre los dos ligandos, $P_{previo}(F_{ji})$, y después de realizar el cálculo de FEP ajusta la distribución basándose en el resultado calculado de FEP (la distribución posterior $P_{pos}(F_{ji} | E_{ji})$).

Antes de realizar el cálculo de FEP, si no se conoce información sobre el cálculo de energía libre de unión relativa, una suposición razonable sobre la distribución previa de F_{ji} sería que F_{ji} se distribuye uniformemente entre $-\infty$ y $+\infty$, es decir

$$P_{previa}(F_{ji}) = 1 \text{ para } F_{ji} \in [-\infty, +\infty]$$

Obsérvese que la distribución previa no está normalizada.

Si el cálculo de FEP da un valor de E_{ji} , entonces según estadística bayesiana, la distribución posterior de F_{ji} sería:

$$P_{post}(F_{ji} | E_{ji}) = \frac{\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(E_{ji} - F_{ji})^2}{2\sigma^2}\right)}{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(E_{ji} - F_{ji})^2}{2\sigma^2}\right) dF_{ji}} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(E_{ji} - F_{ji})^2}{2\sigma^2}\right)$$

Para el caso descrito anteriormente, usando tres ligandos L_1 , L_2 y L_3 , el método realiza cálculos de FEP que forman un ciclo cerrado, con la diferencia de energía libre calculada E_{21} , E_{32} y E_{13} . Entonces según la estadística bayesiana, la distribución posterior de (F_{21}, F_{32}, F_{13}) es:

$$P_{post}(F_{21}, F_{32}, F_{13} | E_{21}, E_{32}, E_{13}) = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{2/3} \exp\left(-\frac{(E_{21} - F_{21})^2}{2\sigma^2} - \frac{(E_{32} - F_{32})^2}{2\sigma^2} - \frac{(E_{13} - F_{13})^2}{2\sigma^2}\right)}{\int_{-\infty}^{+\infty} dF_{21} \int_{-\infty}^{+\infty} dF_{32} \int_{-\infty}^{+\infty} dF_{13} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{2/3} \exp\left(-\frac{(E_{21} - F_{21})^2}{2\sigma^2} - \frac{(E_{32} - F_{32})^2}{2\sigma^2} - \frac{(E_{13} - F_{13})^2}{2\sigma^2}\right)}$$

$$P_{post}(F_{21}, F_{32}, F_{13} | E_{21}, E_{32}, E_{13}) \propto \exp\left(-\frac{(E_{21} - F_{21})^2}{2\sigma^2} - \frac{(E_{32} - F_{32})^2}{2\sigma^2} - \frac{(E_{13} - F_{13})^2}{2\sigma^2}\right)$$

con la limitación

$$F_{21} + F_{32} + F_{13} = 0$$

El pico de la distribución posterior se encuentra en

$$\hat{F}_{21} = E_{21} - \frac{1}{3}(E_{21} + E_{32} + E_{13}) = \frac{2}{3}E_{21} - \frac{1}{3}(E_{32} + E_{13})$$

$$\hat{F}_{32} = E_{32} - \frac{1}{3}(E_{21} + E_{32} + E_{13}) = \frac{2}{3}E_{32} - \frac{1}{3}(E_{21} + E_{13})$$

$$\hat{F}_{13} = E_{13} - \frac{1}{3}(E_{21} + E_{32} + E_{13}) = \frac{2}{3}E_{13} - \frac{1}{3}(E_{32} + E_{21})$$

Esto confirma los resultados obtenidos del método de máxima verosimilitud presentado anteriormente.

II. Programación de una implementación general de la invención de estimación de energía libre derivada del cierre de ciclo y de estimación de error

Para un conjunto de ligandos L, el método realiza N cálculos de energía libre relativa que conectan todos los ligandos con cierre(s) de ciclo. La realización particular de la invención descrita en esta sección da las mejores predicciones de las energías libres relativas entre los ligandos y las barras de error asociadas a las predicciones. La entrada es las N diferencias de energía libre calculadas entre los N pares de ligando con simulaciones de energía libre relativa realizadas, y la salida es las N diferencias de energía libre relativa predichas y las barras de error asociadas entre los N pares de ligandos para los que se realizaron las simulaciones de energía libre relativa, y también las diferencias de energía libre relativa entre cualquier par de ligandos en el conjunto y sus barras de error asociadas. Las energías libres relativas predichas no tienen histéresis, que significa que la diferencia de energía libre relativa predicha entre cualesquiera dos ligandos es independiente de las rutas que los conectan, y así son internamente coherentes.

En las rutas de ejemplo representadas en la Figura 3, existen 11 cálculos de energía libre relativa para un conjunto de 8 ligandos que conectan los 8 ligandos con varios cierres de ciclo. La entrada del programa es los 11 resultados de cálculos de energía libre relativa para los 11 pares de ligandos, y la salida es las diferencias de energía libre relativa predicha para los 11 pares del ligandos, sus barras de error asociadas y también las diferencias de energía libre relativa entre cualquier par arbitrario de 8 ligandos en el conjunto y sus barras de error asociadas.

Entrada:

$\{E_i; L_j \rightarrow L_k\}$ donde $i \in \{1, 2, 3, \dots, N\}$

Salida:

$\{F_i, e_i; L_j \rightarrow L_k\}$ donde $i \in \{1, 2, 3, \dots, N\}$

y $\{F_{ij}, e_{ij}\}$ donde $i, j \in \{1, 2, \dots, L\}$

Algoritmo

Etapla 1: Listar las energías libres calculadas para todos los cálculos de energía libre de unión relativa realizados, y sus ligandos inicial y final correspondientes.

Lo que se consigue de esta etapa es una lista:

$$5 \quad \{E_i; L_j \rightarrow L_k\} \text{ donde } i \in \{1, 2, 3, \dots, N\}$$

Aquí E_i es la energía libre calculada de la simulación de energía libre relativa para el i-ésimo cálculo de energía libre de unión relativa. El ligando inicial y final para este cálculo de energía libre de unión relativa es L_j y L_k . Esto es la entrada del programa.

10 Etapla 2: Listar todos los ciclos cerrados en las rutas de cálculo de energía libre de unión relativa y construir la matriz de conectividad del cierre de ciclo (C3M) para todo el gráfico.

En esta etapa, el método busca en todos los pares de ligandos y encuentra todos los pares que forman un ciclo cerrado y convierte esta información en la matriz de conectividad.

15 Por ejemplo, en las rutas de pares de ligandos mostradas anteriormente, los pares 1, 2 y 5 forman un ciclo cerrado. Puesto que la energía libre es una función del estado, la suma de la energía libre a lo largo de dicho ciclo cerrado debe ser 0. En otras palabras

$$F_1 + F_2 - F_5 = 0$$

En el caso general, la limitación de que la energía libre a lo largo de un ciclo cerrado se pueda escribir en la siguiente función de limitación general:

$$\sum_{i=1}^N M_i F_i = 0$$

20 donde F_i es la diferencia de energía libre entre los dos ligandos en los dos vértices de la arista a lo largo del ciclo cerrado, y M_i es 1 si la arista está en la misma dirección que el ciclo cerrado, -1 si la arista está en la dirección opuesta del ciclo cerrado, y 0 si la arista no está en el ciclo cerrado.

En esta etapa, el método lista todos los ciclos cerrados en las rutas de pares de ligandos, y todas las limitaciones. En el ejemplo mostrado anteriormente, los presentes inventores tienen las siguientes limitaciones para los ciclos cerrados:

$$25 \quad F_1 + F_2 - F_5 = 0$$

$$F_3 + F_4 - F_5 = 0$$

$$F_4 - F_7 - F_6 = 0$$

$$F_1 + F_2 - F_4 - F_3 = 0$$

$$F_5 - F_7 - F_6 - F_3 = 0$$

$$30 \quad F_1 + F_2 - F_7 - F_6 - F_3 = 0$$

$$F_9 + F_{11} - F_{10} = 0$$

En el caso general, suponiendo que existe un número total de M ciclos cerrados en las rutas de pares de ligandos, entonces los presentes inventores tienen M ecuaciones de limitación:

$$\sum_{i=1}^N M_{ji} F_i = 0 \text{ donde } j \in \{1, 2, \dots, M\}$$

35 $M_{ji} = M_G$ es una matriz de $M \times N$ y cada fila de la matriz representa una ecuación de limitación para un ciclo cerrado. Los presentes inventores usan M_G para indicar la matriz de conectividad de cierre de ciclo (C3M) de todo el gráfico.

En el ejemplo mostrado anteriormente,

$$M_G = \begin{bmatrix} 1 & 1 & & & & & & & & & -1 \\ & & 1 & 1 & & & & & & & -1 \\ & & & 1 & & & -1 & -1 & & & \\ 1 & 1 & -1 & -1 & & & & & & & \\ & & -1 & & 1 & -1 & -1 & & & & \\ 1 & 1 & -1 & & & -1 & -1 & & & & \\ & & & & & & & 1 & -1 & 1 & \end{bmatrix}$$

Etapas 3: Descomponer todo el gráfico en subgráficos independientes.

En el gráfico de cálculo de energía libre relativa, si no existe ciclo cerrado que conecte dos subgráficos, entonces las energías libres relativas predichas entre un subconjunto de ligandos son independientes de las energías libres relativas predichas entre el otro subconjunto de ligandos. En el ejemplo mostrado anteriormente, las energías libres relativas predichas entre el conjunto de ligandos [1,2,3,4,5] son independiente de las energías libres relativas entre el conjunto de ligandos [6,7,8] debido a que no existe ciclo cerrado que conecte los dos subgráficos. En esta etapa, el gráfico completo se descompone en cada subgráfico independiente, y entonces cada subgráfico independiente se resuelve uno detrás del otro.

Etapas 3a: Revisar cada fila de M_G , registrar el número de columna de los elementos distintos de cero en cada fila y poner el número de columna en una lista

$$L_j = \{n_j; M_{nj} \neq 0\} \text{ donde } j \in \{1, 2, \dots, M\}$$

En esta etapa, el método produce M listas. Cada lista contiene el número de columna de los elementos distintos de cero de cada fila de la matriz C3M. Para dejarlo claro para la siguiente descripción, el conjunto que contiene las M listas generadas en esta etapa, la lista en bruto se denomina el conjunto S_r .

En el ejemplo mostrado anteriormente, esta etapa del método produce un conjunto de listas en bruto que contiene 7 listas

$$L_1 = \{1, 2, 5\}$$

$$L_2 = \{3, 4, 5\}$$

$$L_3 = \{4, 6, 7\}$$

$$L_4 = \{1, 2, 3, 4\}$$

$$L_5 = \{3, 5, 6, 7\}$$

$$L_6 = \{1, 2, 3, 6, 7\}$$

$$L_7 = \{9, 10, 11\}$$

Etapas 3b: Iterar a través de la segunda lista hasta la última lista en el conjunto de listas en bruto S_r , y comparar los elementos contenidos en L_j donde $j \in \{2, 3, \dots, m\}$ con los elementos en la primera lista L_1 . Si contienen un elemento común (o algunos elementos), entonces fusionar todos los elementos en L_j en la primera lista L_1 , y borrar L_j del conjunto de listas en bruto; si no tiene ningún elemento común, mantener la lista L_j en el conjunto de listas en bruto.

En el ejemplo mostrado anteriormente, L_2 , L_4 , L_5 , L_6 contienen elemento(s) común (comunes) con L_1 , por lo que las listas se borran en esta etapa, y sus elementos se fusionan en L_1 . Lo que queda en el conjunto de listas en bruto de esta etapa es

$$L_1 = \{1, 2, 3, 4, 5, 6, 7\}$$

$$L_3 = \{4, 6, 7\}$$

$$L_7 = \{9, 10, 11\}$$

Etapas 3c: Repetir la etapa 3b para las listas que no se han borrado en la etapa 3b y fusionar los elementos en L_1 si contienen elementos comunes con L_1 . Repetir esta etapa hasta que no se puedan borrar más listas.

En el ejemplo mostrado anteriormente, lo que queda en el conjunto de listas en bruto S_r de esta etapa es

$$L_1 = \{1,2,3,4,5,6,7\}$$

$$L_7 = \{9,10,11\}$$

- 5 Etapa 3d: Mover la primera lista L_1 en el conjunto de listas en bruto S_r en el conjunto de listas final $S_f = \{L_1, \dots\}$ y borrarlo del conjunto de listas en bruto.

En el ejemplo mostrado anteriormente, después de esta etapa en el conjunto de listas final S

$$S_f = \{L_1 = \{1,2,3,4,5,6,7\}\},$$

y queda una lista en el conjunto de listas en bruto

$$S_r = \{L_7 = \{9,10,11\}\}$$

- 10 Etapa 3e: Después de la etapa 3d, si queda más de una lista en el conjunto de listas en bruto, repetir la etapa 3b a etapa 3d hasta que solo quede una lista.

En el ejemplo mostrado anteriormente, no se necesita realizar ninguna acción en esta etapa, puesto que solo queda una lista en el conjunto de listas en bruto después de la etapa 3d.

- 15 Etapa 3f: Mover la última lista en el conjunto de listas en bruto S_r en el conjunto de listas final S_f . Después de esta etapa, quedan varias listas en el conjunto de listas final, y no queda lista en el conjunto de listas en bruto. Las listas en los conjuntos de listas finales no contienen elementos comunes.

En el ejemplo mostrado anteriormente, en el conjunto de listas final los presentes inventores tienen

$$S_f = \{L_1 = \{1,2,3,4,5,6,7\}, L_2 = \{9,10,11\}\}$$

- 20 Etapa 3g: Iterar a través del número de columna 1 a N. Si el número de columna está en cualquier lista en el conjunto de listas final, saltarlo, de otro modo, añadir una lista que solo contenga ese número de columna en el conjunto de listas final.

Esta es la etapa final en este proceso. Después de esta etapa, el conjunto de listas final contiene varias listas, y no existe elemento común entre ninguna de las listas. El número de columna 1 a N está en solo una de las listas en el conjunto de listas final.

- 25 En el ejemplo mostrado anteriormente, después de esta etapa:

$$S_f = \{L_1 = \{1,2,3,4,5,6,7\}, L_2 = \{9,10,11\}, L_3 = \{8\}\}$$

Después de este proceso, todo el gráfico se descompone en cada subgráfico independiente individual. Cada subgráfico independiente contiene las aristas con los números de aristas listados en cada lista del conjunto de listas final S_f . Estos subgráficos independientes no tienen una arista compartida, pero pueden tener vértices compartidos.

- 30 En el ejemplo mostrado anteriormente, todo el gráfico se descompone en 3 subgráficos independientes. El primer subgráfico contiene 7 rutas de cálculo de energía libre de unión [E1, E2, E3, E4, E5, E6, E7], el segundo subgráfico contiene 3 rutas de cálculo de energía libre de unión relativa [E9, E10, E11] y el tercer subgráfico contiene 1 ruta de cálculo de energía libre de unión relativa [E8]. Estos subgráficos se muestran en la Figura 4.

Etapa 4: Conseguir la C3M para cada subgráfico independiente M_s .

- 35 De la etapa 3, todo el gráfico de conectividad ya está descompuesto en subgráficos independientes, y los números de arista de cada subgráfico se almacenan en la lista en el conjunto de listas final. En esta etapa, se logra conseguir C3M para cada subgráfico independiente.

Coger cada lista L_i del conjunto de listas final S_f , se supone que existen N_i elementos en la lista. Buscar en todas las filas de C3M de todo el gráfico, M_G . Para cada fila de M_G , buscar todos los n elementos cuyos números de columna estén en la lista L_i , y si alguno de ellos no es cero, poner los n elementos de columna de esta fila dentro de M_s^i , la C3M para el subgráfico correspondiente a la lista L_i . Hacer esto para todas las filas de M_G y para todas las listas en el conjunto de listas final S_f .

- 40 A partir de esta etapa, C3M de todo el grupo, M_G , una matriz $M \times N$, se descompone en cada C3M independiente, M_s^i , una matriz $M_i = N_i$, cada una correspondiente a un subgráfico independiente. La dimensionalidad de estas matrices cumple las siguientes propiedades:
- 45

$$\sum_i M_i = M$$

$$\sum_i N_i = N$$

Después de esta etapa, todo el gráfico se descompone en subgráficos independientes, y la C3M de todo el gráfico, M_G , también se descompone en C3M para cada subgráfico independiente, M_S^i .

- 5 Obsérvese que cada fila en M_S^i representa un ciclo cerrado en la ruta, y la matriz de conectividad M_S^i para algunos subgráficos independientes puede contener cero elementos, que significa que no existe ciclo cerrado en ese subgráfico.

En el ejemplo anterior, las C3M para los tres subgráficos independientes son:

$$M_S^1 = \begin{bmatrix} 1 & 1 & & & -1 \\ & & 1 & 1 & -1 \\ & & & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & & \\ 1 & 1 & -1 & & -1 & -1 \end{bmatrix}$$

$$M_S^2 = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}$$

$$M_S^3 = \emptyset$$

Obsérvese que M_S para el tercer subgráfico está vacío.

Etapas 5: Reducir C3M para cada subgráfico M_S en la matriz independiente M_{SR}

- 15 En la C3M generada a partir de la etapa 4 para cada subgráfico, algunas filas no son independientes entre sí, que significa que los ciclos cerrados listados en M_S no son independientes. Por ejemplo, en el ejemplo mostrado anteriormente, para el subgráfico 1, existen tres ciclos independientes, pero los seis ciclos cerrados se enumeran en la matriz de conectividad M_S^1 . Tres de ellos no son necesarios.

- 20 En esta etapa, la C3M de cada subgráfico, M_S^i , se reduce en la matriz que solo contiene vectores de fila independientes. Los presentes inventores lo llaman la matriz reducida que solo contiene vectores de fila independientes, RC3M (la matriz de conectividad de cierre de ciclo reducida), y lo indican como M_{SR}^i .

Existen muchas formas de reducir la matriz M_S^i en M_{SR}^i . Aquí, se lista una forma simple.

Etapas 5a: Poner las primeras dos filas de M_S^i para formar una nueva matriz M de $2 \times N_i$.

Etapas 5b: Calcular el determinante $|MM^T|$. Si el determinante no es cero, entonces las dos filas son independientes; si el determinante es cero, entonces las dos filas son dependientes, y borrar la segunda fila.

- 25 Etapas 5c: Poner la tercera fila de M_S^i en M, y repetir la etapa 5b.

Etapas 5d: Repetir la etapa 5c para todas las filas restantes en M_S^i .

Después de esta etapa, se obtiene M_{SR}^i para cada subgráfico. M_{SR}^i solo contiene vectores de filas independientes, que significa que solo se enumeran ciclos cerrados independientes en la matriz.

En el ejemplo anterior, M_{SR} para el primer subgráfico es:

$$M_{SR}^1 = \begin{bmatrix} 1 & 1 & & -1 \\ & & 1 & 1 & -1 \\ & & & 1 & & -1 & -1 \end{bmatrix}$$

Obsérvese que, dependiendo del orden de las filas en C3M y el algoritmo para reducir la matriz, M_{SR} para cada subgráfico no es único, pero el número de filas en M_{SR} es independiente del orden de filas y del algoritmo para hacer la reducción.

Después de esta etapa, se obtiene la matriz de conectividad de cierre de ciclo reducida M_{SR} para cada subgráfico independiente. M_{SR} solo contiene ciclos independientes en cada subgráfico. A continuación, se resolverá la ecuación de cierre de ciclo para cada subgráfico independiente.

Etapla 6: Resolver la ecuación de cierre de ciclo para cada subgráfico independiente.

Esta etapa produce, para cada subgráfico, un conjunto de estimaciones óptimas de energías libres para los pares N_i de ligandos sin histéresis, y también maximiza la probabilidad.

Existen dos casos diferentes.

Caso a: No existe ciclo cerrado en el subgráfico. Entonces la solución a este caso es:

$$F_i = E_i$$

$$e_i = 0$$

En el ejemplo mostrado anteriormente, el tercer subgráfico no tiene ningún ciclo cerrado. Así la solución para el tercer subgráfico es

$$F_8 = E_8$$

$$e_8 = 0$$

Caso b: Existen m ciclos cerrados independientes en el subgráfico.

En este caso, para el subgráfico correspondiente a las listas L_s en el conjunto de listas final S_i , con n elementos y la RC3M correspondiente, $M = M_{SR}$ (matriz $m \times n$), se minimiza la siguiente función:

$$\sum_i^n (F_i - E_i)^2 \text{ donde } i \in L_s$$

con las limitaciones:

$$\sum_i M_{ji} F_i = 0 \text{ donde } j \in \{1, 2, \dots, m\}, i \in L_s$$

Usando el multiplicador de Lagrange, los presentes inventores pueden necesitar minimizar la siguiente función:

$$L = \sum_i^n (F_i - E_i)^2 - \sum_j^m 2C_j \left(\sum_i M_{ji} F_i \right)$$

con las limitaciones

$$\sum_i M_{ji} F_i = 0 \text{ donde } j \in \{1, 2, \dots, m\}, \text{ y } i \in L_s$$

donde C_j son los coeficientes de los multiplicadores de Lagrange.

Resolviendo la ecuación

$$\frac{\partial L}{\partial F_i} = 2(F_i - E_i) - \sum_j^m 2C_j M_{ji} = 0$$

el método establece:

$$\begin{aligned} F_i &= E_i + \sum_j^m C_j M_{ji} \\ &= E_i + \sum_j^m M_{ji} C_j \\ &= E_i + \sum_j^m M_{ij}^T C_j \end{aligned}$$

Aquí M_{ji}^T es la transpuesta de M_{ji} .

Escrito en la formulación de matriz, donde

$$\vec{E} = \begin{Bmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{Bmatrix}, \vec{F} = \begin{Bmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{Bmatrix}, \vec{C} = \begin{Bmatrix} C_1 \\ C_2 \\ \vdots \\ C_m \end{Bmatrix}$$

5

y

$$M = \begin{bmatrix} M_{11} & M_{12} & \cdot & \cdot & \cdot & M_{1n} \\ M_{21} & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ M_{m1} & M_{m2} & \cdot & \cdot & \cdot & M_{mn} \end{bmatrix}$$

da los siguientes dos conjuntos de ecuaciones:

$$(1): \vec{F} = \vec{E} + M^T \vec{C}$$

$$(2): M \vec{F} = \vec{0}$$

10

Introduciendo la ecuación (1) en la ecuación (2), se consigue la siguiente ecuación:

$$(3): M M^T \vec{C} = -M \vec{E}$$

Tanto la matriz M RC3M como el vector E son conocidos, por lo que la ecuación se puede resolver para lograr los vector de coeficientes C . Después de que se resuelvan los coeficientes, la ecuación (1) da los mejores estimadores de energía libre para los n pares de simulaciones.

15

Etapas 7: Lograr los estimados de error para cada ruta de cálculo de energía libre de unión relativa posible

Según la hipótesis introducida anteriormente, la suma de los cálculos de energía libre relativa a lo largo de un ciclo cerrado se distribuye de forma promedio con la desviación estándar $\sqrt{ns}$ donde n es el número de miembros en el

ciclo cerrado. El error para cada diferencia calculada entre un par de ligandos en este ciclo cerrado es $\frac{\Delta}{\sqrt{n}}$, donde Δ es la histéresis del ciclo cerrado.

20

Para cada par de ligandos en el ciclo cerrado, el error para la diferencia de energía libre calculada entre el par de ligandos en el ciclo cerrado se puede estimar por la histéresis de ese ciclo.

Escrito en una forma más reducida, para cada fila de C3M, M_s (obsérvese que aquí es la matriz de conectividad de cierre de ciclo no reducida), se puede asignar la siguiente estimación de error a cada miembro en el ciclo:

$$(4): e_{ji} = |M_{ji}| \frac{\sum_i^n E_i}{\sqrt{\sum_i^n |M_{ji}|}}$$

- 5 Después de hacer el cálculo de error según la ecuación (4) para todos los ciclos cerrados, que es para todas las filas en M_s , la estimación de error final para la estimación de energía libre de unión relativa i se elige como el máximo de las estimaciones de todas las filas. En otras palabras:

$$(5): e_i = \text{Máx}\{e_{ji}; j \in \{1, 2, \dots, m\}\} = \text{Máx}\{|M_{ji}| \frac{\sum_i^n E_i}{\sqrt{\sum_i^n |M_{ji}|}}; j \in \{1, 2, \dots, m\}\}$$

Después de la etapa 6 y la etapa 7, se consiguen las mejores estimaciones de energía libre para las n energías libres relativas y también sus barras de error asociadas para cada subgráfico independiente.

- 10 Etapa 8: Informar la estimación de la diferencia de energía libre entre cualquier par de ligandos en el conjunto y sus barras de error asociadas

En las etapas previas, el método ya ha proporcionado las mejores estimaciones de energía libre y sus barras de error asociadas para los n pares de ligandos en cada subgráfico independiente. La fusión de estos datos de todos los subgráficos independientes produce la salida:

$$\vec{F}_i = \begin{Bmatrix} F_1 \\ F_2 \\ \cdot \\ \cdot \\ F_N \end{Bmatrix} \quad \text{y} \quad \vec{e} = \begin{Bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_N \end{Bmatrix}$$

- 15 Para cualquier par de ligandos que no están listados en lo anterior, el método lista todas las posibles rutas en el gráfico que pueden conectar los dos ligandos, y entonces suman las energías libres en cada arista de las rutas y también los errores. La energía libre es independiente de la ruta, pero los errores pueden depender de la ruta. El máximo de todos los errores a informar se puede escribir en una fórmula concreta:

$$F_{ij} = \sum_k I_k F_k$$

$$e_{ij} = \text{Máx}\{e_{ij}^k = \sqrt{\sum_k e_k^2}\}$$

donde

$$I_k = \begin{cases} 1; & \text{cuando } F_k \text{ está en la misma dirección que la ruta} \\ -1; & \text{cuando } F_k \text{ está en la dirección opuesta a la ruta} \end{cases}$$

- 25 y se coge el máximo para todas las posibles rutas que conectan el ligando L_i y el ligando L_j . Obsérvese que en esta etapa, las estimaciones de energía libre y los errores solo se calculan para los pares de ligandos donde no se realizan los cálculos de energía libre (no directamente conectados en el gráfico). Las estimaciones de energía libre y los errores para los pares de ligandos donde se realizan los cálculos de energía libre se calculan en etapas previas.

- 30 Para un experto en la técnica, debe ser evidente en lo anterior que se pueden adoptar muchos enfoques diferentes para desarrollar una implementación informática de un modelo probabilístico de las probables diferencias de energía libre y sus errores que dan lugar a la(s) histéresis observada(s) para ciclo(s) cerrado(s); y para hacer predicciones óptimas de las diferencias de energía libre y sus errores dado dicho modelo probabilístico. Por ejemplo, sin limitación, se pueden usar métodos de matrices o de gráficos para crear la estructura de datos requerida, se pueden usar métodos algebraicos matriciales o de teoría de gráficos para construir el modelo probabilístico, y se pueden usar métodos de máxima verosimilitud o bayesianos para parametrizar el modelo para que esté de acuerdo con la(s) histéresis. La formulación algebraica matricial en la sección precedente es solo una realización particular de la invención, y puede

ser fácilmente generalizada de varias formas, que incluye suponer que el error más probable de las ramas de la vía se suma al error de Bennett, el error de muestreo con reposición, o ambos. Asimismo, se podría generalizar para usar otras distribuciones aparte de la distribución gaussiana, tal como, sin limitación, las distribuciones de Lorentz, Levy o Raileigh, delta y uniformes.

- 5 Como tal, la variedad de enfoques que se podrían considerar para construir un modelo probabilístico particular de las probables diferencias de energía libre y sus errores que dan lugar a la(s) histéresis observada(s) para ciclo(s) cerrado(s); y para hacer predicciones óptimas de las diferencias de energía libre y sus errores dado dicho modelo probabilístico serían realizaciones interesantes de la invención descrita en el presente documento.

- 10 Ejemplos ilustrativos de aplicaciones de la invención a diversos cálculos de energía libre de unión relativa particulares a diversos receptores y ligandos particulares

Algoritmo de cálculo de energía libre relativa de FEP/REST

- 15 La invención mejora la exactitud y la fiabilidad de los cálculos de energía libre relativa (por ejemplo, cálculos explícitos del modelo de disolvente). Dicha exactitud es particularmente difícil de lograr si los restos de proteína que rodean el sitio de unión se mueven significativamente cuando se unen ligandos diferentes, o los dos ligandos adoptan diferentes modos de unión o manifiestan uno o más modos de unión degenerados.^{4,11,19} En tales casos, el muestreo de todas las conformaciones relevantes mejora la exactitud de las estimaciones de la diferencia de energía libre, incluso los tiempos de transición hasta el intercambio entre las diferentes conformaciones pueden ser demasiado grandes para estar accesibles en la escala de tiempo de la mayoría de las longitudes de simulación de FEP/MD.

- 20 Aquí, los presentes inventores presentan un conjunto particular de cálculos de energía libre de unión relativa que ilustra la utilidad del método usando un método de cálculo de energía libre relativa recientemente desarrollado, FEP/REST. FEP/REST, introducido por Wang et al.,¹² incorpora el método de muestreo mejorado, REST, en los cálculos de FEP mediante un eficiente protocolo de salto λ , para muestrear transposiciones estructurales locales relevantes en los cálculos relativos de afinidad de unión proteína-ligando dentro de tiempos de simulación fácilmente accesibles. Los presentes inventores también presentan cálculos de energía libre relativa de FEP/MD más rutinarios, para ilustrar similarmente la utilidad de la invención, independiente de la metodología de cálculo de energía libre de unión relativa elegida particular.

- 25 FEP implica calcular la diferencia de energía libre entre dos sistemas por transformación alquímica entre los sistemas en una serie de etapas discretas representadas por valores λ , donde λ varía desde 0 para el estado inicial hasta 1 para el estado final. FEP/REST modifica la energía potencial para una región localizada que rodea el sitio de unión (que los presentes inventores llaman la región 'caliente') que puede incluir restos de proteína, así como el ligando. Para las ventanas intermedias de λ , la energía potencial para la región localizada se cambia de escala en un factor inferior a 1. De esta forma, se reducen las barreras de energía, que permite el eficiente muestreo de las diferentes conformaciones en estas ventanas λ intermedias, y las diferentes conformaciones se propagan a los estados finales mediante el método de intercambio de réplicas hamiltoniano.^{15,22,23} Los presentes inventores llaman región 'caliente' a la región donde se cambia de escala la energía potencial, puesto que las contribuciones locales a la energía potencial han cambiado de escala hasta valores más pequeños de forma que la temperatura eficaz de esa región sea más alta, aunque esta analogía no se traslada a las partículas que en realidad tienen energías cinéticas más altas en la simulación. Los detalles sobre el algoritmo de FEP/REST se describen en el trabajo de FEP/REST por Wang et al.¹².

- 30 La región 'caliente' para el muestreo mejorado por REST se configura aquí para incluir el grupo funcional del ligando mutado en los estudios de SAR. Si el grupo funcional del ligando está unido a un anillo aromático, el anillo aromático también se incluye en la región 'caliente', que permite que el anillo se dé la vuelta mediante el muestreo mejorado de REST. El perfil de temperatura eficaz para la región 'caliente' de REST se calcula según la ecuación 9 de la referencia 26²⁴, estableciendo la relación de aceptación esperada de intercambio de réplicas para que sea 0,3, un valor óptimo que refleja las compensaciones entre el número de réplicas necesarias y la eficiencia de intercambio. Se usa un total de 12 ventanas λ para tanto los cálculos de FEP normal como de FEP/REST. Se desactivaron las interacciones electrostáticas que son únicas para el estado inicial antes de las interacciones de Lennard-Jones (LJ), y se activaron las interacciones de LJ que son únicas para el estado final antes de las interacciones electrostáticas. Las interacciones unidas se convirtieron suavemente desde el estado inicial hasta el estado final. Se suavizó el núcleo de las interacciones LJ para evitar las singularidades e inestabilidades en las simulaciones.²⁵ En nuestras presentes simulaciones, el anillo de benceno en el que se variaron los sustituyentes en los estudios de SAR¹⁶ están incluidos en la región 'caliente', y la temperatura eficaz más alta con un total de 12 ventanas λ alcanzada es aproximadamente 900K (626,85 °C).

Detalles de los diversos cálculos de energía libre relativa de FEP/REST y FEP/MD:

- 35 Se usó el módulo de mutación del grupo funcional de ligando implementado en Desmond^{26,27} para configurar los cálculos.²⁸ Se usó el campo de fuerza OPLS2005^{29,30} para las proteínas y los ligandos junto con el modelo de agua de SPC³¹. Las estructuras iniciales para las simulaciones se tomaron de las estructuras de PDB con ID 1h1q, 1h1r, 1h1s, 1oiu, 1oi9 y 1oiy.¹⁶ Siempre que estuvieron presentes múltiples conformaciones en PDB, los presentes inventores seleccionaron la primera conformación para su uso en las simulaciones. La proteína se preparó usando

Wizard de preparación de proteínas^{32,33} durante el que se asignaron los estados de protonación suponiendo un pH de 7,0.³⁴ Se retuvo la ciclina A con CDK2 y se trató como parte del receptor para todos los cálculos. Se añadieron iones Na⁺ o Cl⁻ para mantener la neutralidad eléctrica. Se relajaron los sistemas y se equilibraron usando el protocolo de relajación de Desmond por defecto, que implica una serie de minimizaciones y simulaciones de dinámica molecular corta. Se usaron un total de 12 ventanas λ para tanto los cálculos de FEP normales como de FEP/REST. La etapa de producción duró 2 ns para las simulaciones complejas y 5 ns para las simulaciones de disolvente. Se intentaron intercambios de réplicas entre ventanas lambda vecinas cada 1,2 ps. Las etapas de producción se muestrearon usando tanto FEP/MD como FEP/REST. Se usó el método de relación de aceptación de Bennett (BAR)³⁵ para calcular la energía libre. Se estimaron los errores para cada cálculo de energía libre usando tanto muestreo con reposición^{19,36} como la predicción de errores analíticos de BAR,^{19,36,37} y se informó el mayor de dos errores.

Evaluación de la coherencia y fiabilidad de las predicciones a modo de una aplicación particular de la invención

Se pueden ordenar jerárquicamente las afinidades de unión relativas para todo el conjunto de ligandos en cálculos de afinidad de unión relativa proteína-ligando por FEP, lo más simple, realizando todas las simulaciones de FEP de un ligando de referencia común. Además, si se obtiene un resultado completamente convergente para cada cálculo, las predicciones finales son, teóricamente, independientes de las rutas elegidas para hacer los cálculos de energía libre de unión relativa. Por ejemplo, para tres ligandos A, B y C, existen dos estrategias para obtener la afinidad de unión relativa entre el ligando A y el ligando C: (1) muestra la ruta directamente desde A hasta C; o (2) muestra la ruta en dos etapas, desde A hasta B y B hasta C, y entonces suma las dos energías libres obtenidas. Teóricamente, la estimación de energía libre convergida final debe ser la misma a partir de los métodos anteriores. En la práctica, sin embargo, debido a los errores en cada cálculo, que incluyen los errores no sesgados debido a las fluctuaciones aleatorias en el muestreo del espacio de fase que se puede caracterizar por el límite de error de Bennett, y el error sesgado debido a un muestreo *sistemáticamente* incompleto del espacio de fase que no se caracterizaría por el límite de error de Bennett, tanto las varianzas como los valores promedio de las dos rutas anteriores son normalmente algo diferentes.³⁸ Dadas estas limitaciones, es importante evaluar la fiabilidad de los resultados obtenidos.

Para conseguir información adicional sobre la coherencia y la fiabilidad de las predicciones, los presentes inventores incluyen cierres de ciclo para todas las rutas. En el ejemplo dado anteriormente, si los resultados de las dos rutas coinciden dentro de un umbral especificado, indica que las predicciones son coherentes, y probablemente fiables dentro de los límites del campo de fuerza empleado. Sin embargo, si los resultados de las dos rutas se diferencian en más de cierto valor umbral, puede indicar que los resultados pueden no converger y las predicciones pueden no ser fiables. El valor umbral depende del número de miembros en cada ciclo y el error que se espera por rama en el ciclo. En esta realización particular de la invención se evalúan rigurosamente la fiabilidad de las predicciones y cómo asignar adaptativamente un valor umbral apropiado para cada ciclo. Además, una vez los presentes inventores incluyen cierres de ciclo en la ruta, existen múltiples formas independientes de calcular la diferencia de energía libre entre dos estados. En esta realización particular de la invención, también se informan predicciones óptimas de las múltiples estimaciones de energía libre a lo largo de los ciclos cerrados. La invención en esta realización también proporciona directamente información sobre el error de convergencia asociado a los cálculos de energía libre de la histéresis de cada ciclo.

Los métodos tratados en el presente documento se realizan por un sistema informático programado para realizar las etapas. En la Figura 15, el sistema incluye ordenador(es) 110, almacenes de datos 120 y módulos de programa 130. Las instrucciones del ordenador se almacenan en uno o más medios legibles por ordenador no transitorios. La Figura 16 muestra las etapas en el método que incluyen un listado de energías libres calculadas para todas las rutas de cálculo de energía libre (302), listado de todos los ciclos cerrados y construcción de la matriz de conectividad de cierre de ciclo (304) y descomposición del gráfico completo en subgráficos independientes (306).

A continuación, los presentes inventores tratan ciertos ejemplos específicos y presentan tablas relacionadas del siguiente modo:

Tabla 1: Las energías libres de unión relativa predichas por FEP/MD y FEP/REST para los cinco ligandos con estructuras cristalinas en comparación con el ligando de referencia 1h1q.

Tabla 2: Las predicciones de energía libre final para los 6 ligandos con estructuras cristalinas de cálculos de FEP/MD y FEP/REST usando la fórmula derivada por las realizaciones de la invención.

Tabla 3: Las estructuras de inhibidores de CDK2 y su compuesto N^o estudiado usando FEP/REST.

Tabla 4: Las energías libres de unión relativas predichas por FEP/REST para los 10 compuestos adicionales sin estructuras cristalinas.

Tabla 5: Las predicciones de energía libre final para todo el conjunto de ligandos usando la fórmula derivada por las realizaciones de la invención.

Tabla 6: Resultados de FEP/REST para las 8 rutas de cálculo de energía libre de unión relativa de los ligandos jnk1.

Tabla 7: Las afinidades de unión predichas y sus errores asociados para los 7 ligandos que se unen a jnk1.

Tabla 8: Resultados de FEP/REST para las 7 rutas de cálculo de energía libre de unión relativa para los ligandos de BACE.

Tabla 9: Las afinidades de unión predichas y sus errores asociados para los 6 ligandos que se unen a BACE.

Tabla 10: Resultados de MC/FEP para las 9 rutas de cálculo de energía libre de unión relativa de ligandos de escitalona deshidratasa.

Tabla 11: Las afinidades de unión predichas y sus errores asociados para los 3 ligandos que se unen a escitalona deshidratasa.

Tabla 12: Las diferencias de energía libre predichas de cierre de ciclo entre los cuatro estados representados en la Figura 14, y sus errores asociados.

Resultados y discusiones

Rendimiento sobre los ligandos con estructuras cristalinas

Se llevaron a cabo simulaciones de FEP/MD y FEP/REST para predecir las afinidades de unión relativas para los seis ligandos presentados en la Figura 5 que se unen al receptor de CDK2-ciclina A. Estos son solo seis ligandos con estructuras cristalinas disponibles para cada holo-complejo, con ID de PDB 1h1q, 1h1r, 1h1s, 1oi9, 1oiu y 1oiy,¹⁶ y los datos de afinidad de unión experimentales para estos ligandos proceden de la misma publicación usando el mismo método. Así, los cálculos informados aquí son esencialmente puramente un ejercicio de puntuación, puesto que los modos de unión de ligandos se conocen de la cristalografía. En estas estructuras cristalinas, la conformación del receptor de CDK2 es esencialmente la misma, independientemente de qué ligando se una. Por tanto, los presentes inventores no esperaron que los cambios conformacionales del muestreo de proteína limitaran significativamente la convergencia de la simulación. En los cálculos, los presentes inventores eligieron usar la conformación de proteína y ligando de 1h1q como la estructura inicial para las cinco perturbaciones (dibujadas con flechas azules en la Figura 5), puesto que todos los otros ligandos en el conjunto se generan añadiendo diferentes sustituyentes en el anillo de benceno del ligando 1h1q.

Comparación entre resultados de FEP/MD y FEP/REST

Los resultados de FEP/MD y de FEP/REST para las rutas de cálculo de energía libre de unión relativa entre el ligando con estructuras cristalinas se dan en la Tabla 1. La correlación de las predicciones con datos de afinidad de unión experimental se presenta en la Figura 6. A pesar del intervalo de ~3,5 kcal/mol de afinidades de unión experimentales, FEP/MD predice que los cinco compuestos (que excluyen el compuesto de referencia 1h1q) tienen casi las mismas afinidades de unión, dando como resultado un valor de R^2 de 0,32. Por comparación, los resultados de FEP/REST se correlacionan altamente con los valores experimentales ($R^2=0,91$), y las desviaciones de los resultados de FEP/REST de los valores experimentales son inferiores a 0,8 kcal/mol para todos los pares de ligandos.

A partir de las comparaciones de los resultados de FEP/MD y de FEP/REST, parece que las predicciones para los compuestos 1oi9, 1oiu y 1oiy de los dos métodos son esencialmente las mismas, y que el principal motivo para la correlación mejorada con los resultados experimentales procede de las mejores estimaciones de la afinidad de unión para los ligandos 1h1s y 1h1r usando FEP/REST. Más adelante los presentes inventores tratarán por qué FEP/REST supera significativamente a FEP/MD para estos dos casos.

Una revisión de las trayectorias de FEP/MD y de FEP/REST para el ligando 1h1r mostró que el anillo de benceno sustituido con 2-cloro se da la vuelta varias veces en la trayectoria de FEP/REST, mientras que el anillo sigue estando atrapado en la conformación inicial en la trayectoria de FEP/MD. La conformación inicial, correspondiente a una de las conformaciones en la estructura cristalina, y la conformación dada la vuelta alternativa, observada en el cálculo de FEP/REST, se presentan en la Figura 7. Es interesante señalar que, para el ligando 1h1r, la densidad electrónica en la estructura cristalina de rayos X depositada también indicó la presencia de un modo de unión a ligando alternativo (Figura 7 panel derecho) que corresponde a la alternativa que se encontró en los cálculos de FEP/REST.

Para 1h1s, además del modo de unión modelado en la estructura cristalina, se muestrea otra conformación en la trayectoria de FEP/REST, pero no en las simulaciones de FEP/MD. La Figura 8 muestra el modo de unión encontrado en la estructura cristalina y el modo de unión alternativo muestreado en el cálculo de FEP/REST. En la estructura cristalina, el grupo sulfonamida forma dos enlaces de hidrógeno (HB) con Asp86: uno con el esqueleto y otro con la cadena lateral. En el modo de unión alternativo muestreado en la simulación de FEP/REST, el grupo sulfonamida gira aproximadamente 120° alrededor del enlace CS con respecto a la estructura cristalina. Los restos de proteína que rodean el sitio de unión, particularmente Gln85 y Asp86, se reordenan en consecuencia y se forman dos enlaces de hidrógeno entre el grupo sulfonamida y los restos de proteína: uno con la cadena lateral de Asp86 y el otro con la cadena lateral de Gln85. Los factores B de estos pocos residuos que rodean el sitio de unión y el ligando en la cristalografía de rayos X indican que estos grupos son bastantes móviles.¹⁶ Por tanto, es bastante posible que el ligando pueda cambiar entre los dos modos de unión en disolución. En la trayectoria de simulación de FEP/REST, la relación de ocupación entre el modo de unión observado cristalográficamente y el modo de unión alternativo es aproximadamente 60:40, en favor del modo cristalográficamente determinado. Sin embargo, el muestreo de este modo

de unión alternativo parece ser muy importante para la correcta predicción de la afinidad de unión, y es responsable de la mayor eficacia de FEP/REST en calcular con exactitud la energía libre de unión relativa de esta especie en comparación con FEP/MD. Es interesante señalar que la invención permite que esto se determine sin referencia a los datos experimentales, puesto que los errores asociados calculados a FEP/MD calculados por esta realización particular de la invención son muy altos (1,01 kcal/mol para 1oiu y 1h1s, y ~0,7 kcal/mol para 1oiy, 1oi9 y 1h1s), donde se calcula por la invención que los errores calculados asociados a FEP/REST son bajos (entre 0,2~0,4 kcal/mol para todos los ligandos).

Evaluación de la coherencia y la fiabilidad de los cálculos de energía libre relativa con una realización particular de la invención

Para evaluar la coherencia y la fiabilidad de las predicciones de FEP/REST, los presentes inventores realizaron tres simulaciones más que forman tres cierres de ciclo en las rutas de cálculo de energía libre de unión relativa (Figura 5). Los resultados de FEP/REST para las tres rutas que forman estos ciclos se dan en las últimas tres filas de la Tabla 1. Las predicciones finales para las afinidades de unión de estos ligandos determinadas a modo de una realización particular de la invención se dan en la Tabla 2. Los ciclos se cierran por sí mismos de forma consistente con discrepancias inferiores a 0,8 kcal/mol de las diferentes rutas para los tres ciclos, que indica que la energía libre converge bien y las predicciones son muy fiables dentro de los límites del campo de fuerza subyacente.

Sin embargo, aún cuando los ciclos estén bien cerrados y las energías libres calculadas a partir de las diferentes rutas coinciden dentro de un valor umbral bastante ajustado, no son exactamente los mismos. Así, para el mismo par de ligandos, los presentes inventores tienen múltiples estimaciones diferentes para las diferencias de energía libre que dependen de cómo se atraviesan las rutas entre el ligando. Entonces se plantea la cuestión de cómo incorporar mejor toda la información que se genera en los cálculos para obtener las mejores estimaciones estadísticas de las energías libres. Los presentes inventores usan una realización particular de la invención para determinar las estimaciones óptimas de las energías libres y los errores asociados a estas predicciones. Aquí, todos los ciclos están bien cerrados, y las predicciones finales con cierres de ciclo coinciden bien con las predicciones anteriores sin cierres de ciclo, validando aún más la alta fiabilidad de estos cálculos. La desviación entre las predicciones finales y los resultados experimentales son todas inferiores a 0,7 kcal/mol y tienen una correlación muy alta con los datos experimentales ($R^2=0,92$).

A diferencia, cuando esta realización particular de la invención se aplicó a ciclos cerrados similares calculados con los cálculos de energía libre relativa de FEP/MD, las histéresis de los ciclos cerrados en los que participa el ligando 1h1r y 1h1s son muy grandes (de hasta el 1,75 kcal/mol y 1,35 kcal/mol, respectivamente), que indica que los cálculos de FEP/MD no convergen. Usando las estimaciones de energía libre de cierre de ciclo a modo de esta realización particular de la invención, los errores para estos cálculos de energía libre de FEP/MD son mucho más grandes (1,01 kcal/mol y 0,68 kcal/mol, respectivamente) que los resultados de FEP/REST (los errores de 0,35 kcal/mol y 0,25 kcal/mol, respectivamente).

Rendimiento sobre los ligandos sin estructura cristalina

Para validar aún más el modelo de los presentes inventores, se realizaron cálculos de FEP/REST adicionales para otros 10 ligandos que se unen al mismo receptor sin estructura cristalina disponible para el holo-complejo. Estos 10 ligandos se seleccionaron aleatoriamente de la misma publicación de Handcastle.¹⁶ Se midieron las afinidades de unión usando el mismo método que para los ligandos anteriormente mencionados con estructuras cristalinas y conservan el mismo núcleo. Las estructuras de los ligandos y sus números de compuesto correspondientes como en la publicación original¹⁶ se dan en la Tabla 3. Las afinidades de unión relativas predichas por FEP/REST para estos ligandos se informan en la Tabla 4 junto con los datos experimentales. Se construyen cierres de ciclo en las rutas para estos ligandos (Figura 9), y las histéresis de los cierres de ciclo son todas inferiores a 1,0 kcal/mol, que indica que los cálculos son convergentes. Las afinidades de unión relativas predichas finales a modo de esta realización particular de la invención se informan en la Tabla 5. Para ligandos con cierres de ciclo construidos en las rutas de cálculo de energía libre de unión relativa, los errores asociados a las predicciones de energía libre también se informan en la Tabla 5. La correlación entre las afinidades de unión relativas predichas y los datos experimentales para todo el conjunto de ligandos (6 ligandos con estructura cristalina y los 10 ligandos sin estructura cristalina) se presenta en la Figura 10. Es evidente que esta realización particular de la invención cuando se aplica a los datos de cálculo de energía libre de FEP/REST rinde bien para estos ligandos sin estructuras cristalinas, que demuestra aún más la transferabilidad del enfoque. La desviación promedio entre las afinidades de unión predichas en comparación con los datos experimentales es 0,44 kcal/mol, y el valor de R^2 es 0,82.

Conclusiones

Anteriormente, los presentes inventores han ilustrado que el método se puede aplicar en múltiples realizaciones al método FEP/REST de cálculo de energía libre recientemente desarrollado, así como al método de cálculo de energía libre FEP/MD más rutinariamente usado para calcular las afinidades de unión relativas entre una serie de ligandos congéneres que se unen al receptor de CDK2-ciclina A. La invención facilitó las comparaciones entre las predicciones de cálculo de energía libre de FEP/REST y FEP/MD. Mientras que el método FEP/MD más rutinario no tuvo buen rendimiento en el orden jerárquico de las afinidades de unión relativas para los seis ligandos con estructura cristalina, la invención proporcionó un valor que indica que el error asociado a las predicciones de FEP/MD era alto. A diferencia, las afinidades de unión relativas calculadas por FEP/REST estuvieron en excelente concordancia con los datos experimentales, y mostraron que convergieron por la aplicación de la invención. La aplicación de la invención a cálculos libres relativos realizados a modo de método de FEP/REST en un conjunto más grande de datos que incluye 10 ligandos sin estructuras cristalinas verificó aún más la capacidad de invención para ordenar jerárquicamente correctamente las afinidades de unión relativas de ligandos congéneres.

La invención proporciona un medio de evaluación de la coherencia y la fiabilidad de los cálculos de energía libre usando información derivada de cálculos del cierre de ciclo asociados a cualquier tipo de procedimiento de cálculo de energía libre relativa. La presente invención proporciona criterios claros por los que juzgar la coherencia y la fiabilidad de las predicciones, así como estimaciones óptimas de las energías libres con límites de error bien formados. La aplicación de la presente invención a la serie de CDK2-ciclina A indica que las predicciones de los presentes inventores usando los cálculos de energía libre relativa de FEP/REST son muy fiables, que también se demuestra por la excelente coincidencia de las predicciones con los datos experimentales. Asimismo, la invención determinó apropiadamente que las predicciones de FEP/MD son menos fiables.

Los presentes inventores incluyen aplicaciones adicionales de diversas realizaciones de la invención a otros sistemas y otro cálculo de energía libre de unión relativa en la generalidad y utilidad de la invención.

Aplicaciones adicionales de diversas realizaciones de la invención a una variedad de sistemas usando una variedad de esquemas de cálculo de energía libre de unión relativa

I. Aplicación de una realización de la invención a la cinasa JNK1

Se realizaron simulaciones de energía libre de unión relativa de FEP/REST para un subconjunto de ligandos que se unen al receptor de jnk1 con dos cierres de ciclo en las rutas de cálculo de energía libre de unión relativa (Fig 11). Los números de ligando para el subconjunto de ligandos son [2,6,7,9,10,12,13], y los 8 resultados de FEP/REST para las 8 rutas de cálculo de energía libre de unión relativa se enumeran en la tabla 6.

Usando los mejores estimadores de energía libre del cierre de ciclo y estimadores de error, las afinidades de unión para los 7 ligandos y sus barras de error asociadas se informan en la Tabla 7.

Obsérvese que las estimaciones de cierre de ciclo pueden no corregir los errores sistemáticos en el campo de fuerza de las simulaciones. Así, las afinidades de unión predichas para dos de los ligandos (ligando 12 y 13) se desvían en más de 1,5 kcal/mol en comparación con los valores experimentales. Las desviaciones pueden ser o debidas al error en el campo de fuerza o en las incertidumbres en las mediciones experimentales. Sin embargo, la correlación global entre los resultados predichos y los valores experimentales es muy buena.

II. Aplicación de una realización de la invención a BACE

Se realizan simulaciones de FEP/REST para un subconjunto de ligandos que se unen al receptor de BACE con dos cierres de ciclo independientes en las rutas de cálculo de energía libre de unión relativa (mostrado en la Figura 12). Los números de ligando para el subconjunto de ligandos son [4j, 4o, 4p, 17d, 17g, 17h], y los resultados de 7 FEP/REST para las 7 rutas de cálculo de energía libre de unión relativa se enumeran en la Tabla 8.

Usando los mejores estimadores de energía libre y estimadores de error del cierre de ciclo, las afinidades de unión para los 6 ligandos y sus barras de error asociadas se informan en la Tabla 9.

Obsérvese nuevamente que la afinidad de unión para el ligando 17h se predice por exceso en aproximadamente 2 kcal/mol, pero la correlación global entre los resultados predichos y los valores experimentales es muy buena.

III. Aplicación de una realización de la invención a cálculos de energía libre de unión relativa de FEP/MC de ligandos que se unen a escitalona deshidratasa

Las afinidades de unión relativas entre 3 ligandos que se unen a escitalona deshidratasa con y sin agua unida en el sitio activo se calculan usando FEP/MC.³⁹ Las rutas de cálculo de energía libre de unión relativa para estas simulaciones con 4 cierres de ciclo independientes se presentan en la Fig 13. Las afinidades de unión calculadas para los cálculos de energía libre de unión relativa se informan en la Tabla 10.

La histéresis del ciclo [L1(H₂O), L2(H₂O), L3(H₂O)] es 4,3 kcal/mol, mucho mayor que 1,4 kcal/mol, que indica que estos cálculos no convergen, y las afinidades de unión calculadas dentro de este ciclo también tienen grandes

desviaciones de los datos experimentales. Similarmente, para el ciclo [L3(H₂O), L3, L2, L2(H₂O)], la histéresis es 1,5 kcal/mol, mayor que 1,4 kcal/mol, que indica que estos cálculos tampoco son fiables.

Por otra parte, las histéresis de los otros ciclos, [L1(H₂O), L1, L3, L3(H₂O)] y [L1, L2, L3], son muy pequeñas (0,2 kcal/mol y 0,5 kcal/mol, respectivamente), que indica que las calculadas son fiables. Las afinidades de unión relativas predichas finales entre los tres ligandos con las estimaciones de cierre de ciclo descritos anteriormente se informan en la Tabla 11.

IV. Aplicación de una realización de la invención para probar el muestreo de energía libre de unión relativa de inserción de partículas para cierres modelo

Se calculan las energías libres de unión relativas entre un metano del átomo unido y los cierres modelo con y sin carga usando el método de inserción de partículas de prueba.⁴⁰ Las energías libres de carga del cierre modelo con y sin el metano unido también se calculan usando el muestreo de FEP/MD. Las rutas de cálculo de energía libre de unión relativa forman un ciclo cerrado (Figura 14) y las energías libres asociadas para cada arista en el ciclo se informan allí.

Usando el análisis de cierre de ciclo descrito anteriormente, la diferencia de energía libre predicha entre los cuatro estados representados en la Fig. 16 y sus errores asociados se informan en la Tabla 12.

TABLAS.

Tabla 1: Energías libres de unión relativas predichas por FEP/MD y FEP/REST para los cinco ligandos con estructuras cristalinas en comparación con el ligando de referencia 1h1q. También se incluyen en la tabla los resultados para las tres rutas de cálculo de energía libre de unión relativa que forman 3 cierres de ciclo para los 6 ligandos con estructuras cristalinas. Los errores para las energías libres calculadas usando estimación del error analítico de BAR también se incluyen en la tabla. Los resultados predichos de FEP/MD tienen grandes desviaciones a partir de los datos experimentales para los ligandos 1h1r y 1h1s (**negrita** en la tabla), mientras que los resultados predichos por FEP/REST coinciden todos bien con los datos experimentales. Se informan las energías libres en unidades de kcal/mol.

Ligando	$\Delta\Delta G(\text{EXP})$	$\Delta\Delta G(\text{FEP/MD})$	$\Delta(\text{FEP/MD-EXP})$	$\Delta\Delta G(\text{FEP/REST})$	$\Delta(\text{FEP/REST-EXP})$
1h1q→Mh1r	0,51	-1,29±0,09	1,80	0,14±0,11	0,37
1h1q→1h1s	-3,07	-1,64±0,27	1,43	-2,72±0,28	0,35
1h1q→1oi9	-1,56	-1,83±0,14	0,27	-1,70±0,16	0,14
1h1q→1oiu	-0,91	-1,68±0,25	0,77	-1,71±0,26	0,80
1h1q→1oiy	-1,61	-1,69±0,22	0,08	-1,92±0,23	0,31
1h1r→1oiu	-1,41	1,36±0,22	2,77	-1,25±0,25	0,16
1oi9→1h1s	-1,51	0,14±0,29	1,65	-0,81±0,30	0,70
1oi9→1oiy	-0,04	1,44±0,24	1,48	0,49±0,24	0,53

Tabla 2: Predicciones de energía libre final para los 6 ligandos con estructuras cristalinas de cálculos de FEP/MD y FEP/REST usando la fórmula derivada por las realizaciones de la invención. También se incluyen en la tabla las estimaciones de error usando la fórmula derivada por realizaciones de la invención para la predicción de energía libre final usando cierre de ciclo. Se usa el ligando 1h1q como referencia, por lo que la desviación para el ligando 1h1q es 0 por diseño. Obsérvese que no se informa de una barra de error de cierre de ciclo para el ligando 1h1q puesto que es el ligando de referencia usado para este trabajo, y su energía libre de unión se establece idénticamente igual a su valor experimental para fines de representación. Las energías libres se informan en unidades de kcal/mol.

	$\Delta G(\text{EXP})$	$\Delta G(\text{FEP/MD})$	$A(\text{FEP/MD-EXP})$	FEP/MD_ Err(cierre_ciclo)	$\Delta G(\text{FEP/REST})$	$\Delta(\text{FEP/REST-EXP})$	FEP/REST_ _Err(cierre_ciclo)
1h1q	-8,18	-8,18	--		-8,18	--	--
1h1r	-7,67	-10,05	2,38	1,01	-8,25	0,58	0,35
1h1s	-11,25	-10,00	1,25	0,68	-10,91	0,34	0,25

	$\Delta G(\text{EXP})$	$\Delta G(\text{FEP}/\text{MD})$	$A(\text{FEP}/\text{MD}-\text{EXP})$	$\text{FEP}/\text{MD}_{\text{Err(cierre_ciclo)}}$	$\Delta G(\text{FEP}/\text{REST})$	$\Delta(\text{FEP}/\text{REST}-\text{EXP})$	$\text{FEP}/\text{REST}_{\text{Err(cierre_ciclo)}}$
1oi9	-9,74	-10,32	0,58	0,75	-10,11	0,37	0,41
1oiu	-9,08	-9,85	0,77	1,01	-9,71	0,63	0,35
1oiy	-9,79	-9,38	0,41	0,75	-9,86	0,07	0,41

Tabla 3: Las estructuras de inhibidores de CDK2 y su compuesto N° estudiado usando FEP/REST.

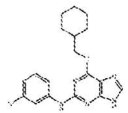
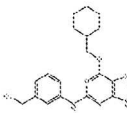
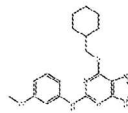
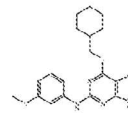
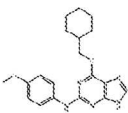
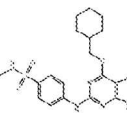
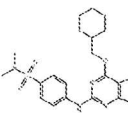
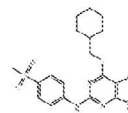
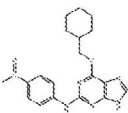
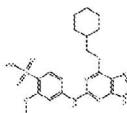
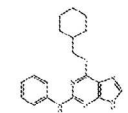
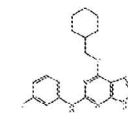
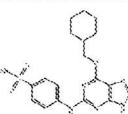
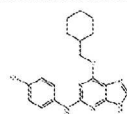
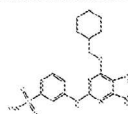
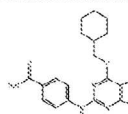
 Compuesto: 17	 Compuesto: 20	 Compuesto: 21	 Compuesto: 22
 Compuesto: 26	 Compuesto: 28	 Compuesto: 29	 Compuesto: 30
 Compuesto: 31	 Compuesto: 32	 Compuesto: 1h1q	 Compuesto: 1h1r
 Compuesto: 1h1s	 Compuesto: 1oi9	 Compuesto: 1oiu	 Compuesto: 1oiy

Tabla 4: Las energías libres de unión relativas predichas por FEP/REST para los 10 compuestos adicionales sin estructuras cristalinas. También se informan en la tabla los errores para las energías libres calculadas usando estimación del error analítico de BAR. Las energías libres se informan en unidades de kcal/mol.

Ligando	$\Delta\Delta G(\text{EXP})$	$\Delta\Delta G(\text{FEP}/\text{REST})$	$\Delta(\text{FEP}/\text{REST}-\text{EXP})$
1h1q→17	1,14	0,19±0,11	0,95
1h1q→20	-0,54	-0,52±0,18	0,02
1h1q→21	0,35	0,64±0,15	0,29
1hq→22	0,42	0,44±0,15	0,02
1h1q→26	-0,25	-1,39±0,14	1,14
1h1q→31	-1,36	-0,57±0,19	0,79
1h1s→28	0,14	0,14±0,20	0,00
1h1s→30	1,44	1,53±0,15	0,09
1h1s→32	1,50	0,78±0,17	0,72
28→29	1,23	1,05±0,18	0,18
1h1r→17	0,63	0,13±0,11	0,50

Ligando	$\Delta\Delta G(\text{EXP})$	$\Delta\Delta G(\text{FEP/REST})$	$\Delta(\text{FEP/REST-EXP})$
21→22	0,07	0,62±0,24	0,55
1h1s→29	1,37	2,19±0,19	0,82
30→29	-0,07	-0,27±0,19	0,20
1oiu→26	0,65	1,99±0,19	1,34

Tabla 5: Las predicciones de energía libre final para el conjunto completo de ligandos usando la fórmula derivada por realizaciones de la invención. También están incluidas en la tabla las estimaciones de error usando las realizaciones de la invención para la predicción de energía final usando cierre de ciclo. Se usa el ligando 1h1q como referencia, por lo que la desviación para el ligando 1h1q es 0 por diseño. Obsérvese que una barra de error de cierre de ciclo no se informa para el ligando 1h1q puesto que es el ligando de referencia usado para este trabajo, y su energía libre de unión se establece idénticamente igual a su valor experimental para fines de representación. Tampoco se informan las barrar de error de cierre de ciclo para los ligandos 20, 31 y 32, ya que aquellas especies fueron técnicamente exigentes de conectar con otros ligandos distintos del ligando de referencia, como se requiere para formar ciclos cerrados para el error. Las energías libres se informan en unidades de kcal/mol.

	$\Delta G(\text{EXP})$	$\Delta G(\text{FEP/REST})$	$\Delta(\text{FEP/REST-EXP})$	Err(cierre_ciclo)
17	-7,04	-8,06	1,02	0,26
20	-8,72	-8,70	0,02	--
21	-7,83	-7,81	0,02	0,47
22	-7,76	-7,47	0,29	0,47
26	-8,43	-8,65	0,22	0,69
28	-11,11	-10,51	0,60	0,58
29	-9,88	-9,20	0,68	0,58
30	-9,81	-9,16	0,65	0,54
31	-9,54	-8,75	0,79	--
32	-9,75	-10,13	0,38	--
R ²	0,82		--	--

Tabla 6: Resultados de FEP/REST para las 8 rutas de cálculo de energía libre de unión relativa de los ligandos jnk1

Ruta	dG_Exp	dG_FEP/REST
Fep2→6	-0,21	-0,72
Fep2→7	-0,18	-0,32
Fep2→9	-0,92	-2,47
Fep2→ 10	-0,59	-0,46
Fep6→13	1,41	0,20
Fep7→ 13	1,38	-0,04
Fep9→ 12	-0,58	-0,88
Fep10→12	-0,91	-2,59

Tabla 7: Las afinidades de unión predichas y sus errores asociados para los 7 ligandos que se unen a jnk1.

Ligando	dG_Exp	dG_CC_pred	Error CC
L2	-8,49	-8,49	0,0
L6	-8,70	-9,17	0,1
L7	-8,67	-8,85	0,1
L9	-9,41	-10,89	0,2
L10	-9,08	-9,03	0,2
L12	-9,99	-11,69	0,2
L13	-7,29	-8,93	0,1

Tabla 8: Resultados de FEP/REST para las 7 rutas de cálculo de energía libre de unión relativa para los ligandos de BACE.

Ruta	dG_Exp	dG_FEP/REST
Fep17d→4j	0,40	0,55
Fep17h→4p	0,25	1,15
Fep17g→4o	0,36	0,63
Fep17d→17h	-0,91	-2,50
Fep17d→17g	-0,32	0,98
Fep4j→4p	-1,06	-2,30
Fep4j→4o	-0,36	-1,00

Tabla 9: Las afinidades de unión predichas y sus errores asociados para los 6 ligandos que se unen a BACE.

Ligando	dG_Exp	dG_CC_pred	Error CC
4j	-9,02	-9,02	0,0
4p	-10,08	-11,22	0,2
4o	-9,38	-10,02	0,2
17d	-9,42	-9,67	0,2
17h	-10,33	-12,27	0,2
17g	-9,74	-10,65	0,2

Tabla 10: Resultados de MC/FEP para las 9 rutas de cálculo de energía libre de unión relativa de ligandos de escitalona deshidratasa.

Ruta	dG_Exp	dG_MC/FEP
L1(H2O)→L1	--	5,5
L2(H2O)→L2	--	-9,8
L3(H2O)→L3	--	-0,4

Ruta	dG_Exp	dG_MC/FEP
L1(H ₂ O)→L3(H ₂ O)	3,8	6,2
L1(H ₂ O)→L2(H ₂ O)	-2,0	14,1
L3(H ₂ O)→L2(H ₂ O)	-5,8	3,6
L1→L3	3,8	0,5
L1→L2	-2,0	-6,3
L3→L2	-5,8	-7,3

Tabla 11: Las afinidades de unión predichas y sus errores asociados para los 3 ligandos que se unen a escitalona deshidratasa. Aún cuando la afinidad de unión predicha para el ligando 3 tenga un gran error relativo, se filtran las rutas de cálculo de energía libre de unión relativa con errores muy grandes y se predice el orden jerárquico correcto.

Ligando	dG_Exp	dG_CC_pred*	Error CC
L1	0	0	0
L2	-1,14	-2,0	0,3
L3	5,82	3,8	0,3
*Nota: las afinidades de unión informadas en la tabla es la afinidad de unión relativa en comparación con el ligando 1.			

5

Tabla 12: Las diferencias de energía libre predichas del cierre de ciclo entre los cuatro estados representados en la Figura 14, y sus errores asociados. M indica el metano del átomo unido, P indica las placas hidrófobas, CP indica las placas cargadas, MP indica las placas hidrófobas unidas con el metano del átomo unido y MCP indica las placas cargadas unidas con el metano del átomo unido.

Mutación	dG_CC_pred	Error CC
M+P→MP	-2,846	0,04
M+CP→MCP	16,540	0,04
P→CP	-131,004	0,04
MP→MCP	-111,618	0,04

10

REFERENCIAS

- (1) Gilson, M. K.; Zhou, H. X. Annu. Rev. Biophys. Biomol. Struct. 2007, 36, 21.
- (2) Jorgensen, W. L. Acc. Chem. Res. 2009, 42, 733.
- (3) Moble, D. L.; Dill, K. A. Structure 2009, 17, 498.
- (4) Gallicchio, E.; Levy, R. M. Curr. Opin. Struct. Biol. 2011, 21, 161.
- (5) Mobley, D. L.; Dill, K. A.; Chodera, J. D. J. Phys. Chem. B 2008, 112, 938.
- (6) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Proc. Nat. Acad. Sci. USA 2007, 104, 813.
- (7) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. J. Am. Chem. Soc. 2008, 130, 2831.
- (8) Wang, L.; Berne, B. J.; Friesner, R. A. Proc. Nat. Acad. Sci. USA 2011, 108, 1326.
- (9) de Amorim, H. L.; Caceres, R. A.; Netz, P. A. Curr. Drug Targets 2008, 9, 1100.
- (10) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. J. Am. Chem. Soc. 2009, 131, 15403.

20

- (11) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. *Curr. Opin. Struct. Biol.* 2011, 21.
- (12) Wang, L.; Berne, B. J.; Friesner, R. A. *Proc. Nat. Acad. Sci. USA* 2012, 109, 1937.
- (13) Huang, X.; Hagen, M.; Kim, B.; Friesner, R. A.; Zhou, R.; Berne, B. J. *J. Phys. Chem. B* 2007, 111, 5405.
- 5 (14) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Nat. Acad. Sci. USA* 2005, 102, 13749.
- (15) Wang, L.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B* 2011, 115, 9431.
- (16) Hardcastle, I. R.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Chen, Y.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Jewsbury, P.; Menyerol, J.; Mesguiche, V.; Newell, D. R.; Noble, M. E.; Pratt, D. J.; Wang, L. Z.; Whitfield, H. J. *J. Med. Chem.* 2004, 47, 3710.
- 10 (17) Malumbres, M.; Barbacid, M. *Nat. Rev. Cancer* 2009, 9, 153.
- (18) Chen, Y. N.; Sharma, S. K.; Ramsey, T. M.; Jiang, L.; Martin, M. S.; Baker, K.; Adams, P. D.; Bair, K. W.; Kaelin, W. G., Jr. *Proc. Nat. Acad. Sci. USA* 1999, 96, 4325.
- (19) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* 2010, 114, 10253.
- 15 (20) Steiner, D.; Oostenbrink, C.; Diederich, F.; Zurcher, M.; van Gunsteren, W. F. *J. Comput. Chem.* 2011, 32, 1801.
- (21) Lai, B.; Oostenbrink, C. *Theor. Chem. Acc.* 2012, 131, 1.
- (22) Jiang, W.; Roux, B. *J. Chem. Theory Comput.* 2010, 6, 2565.
- (23) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* 2003, 107, 13703.
- (24) Patriksson, A.; van der Spoel, D. *Phys. Chem. Chem. Phys.* 2008, 10, 2073.
- 20 (25) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* 1994, 222, 529.
- (26) In Desmond; 3.4 ed.; Schrodinger L. L. C.: New York, NY, 2012.
- (27) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. In *Proceedings of the 2006 ACM/IEEE conference on Supercomputing*; ACM: Tampa, Florida, 2006, p 84.
- 25 (28) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. *J. Chem. Theory Comput.* 2010, 6, 1509.
- (29) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* 1988, 110, 1657.
- (30) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* 2001, 105, 6474.
- 30 (31) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; J., H.; Pullman, B., Ed.; Reidel: Dordrecht, 1981, p 331.
- (32) In Maestro; 9.3 ed.; Schrodinger, L. L. C: New York, NY, 2012.
- (33) In Schrodinger Suite 2012 Protein Preparation Wizard; Schrodinger L. L. C.: New York, NY, 2012.
- 35 (34) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* 2011, 7, 525.
- (35) Bennett, C. H. *J. Comput. Phys.* 1976, 22, 245.
- (36) Paliwal, H.; Shirts, M. R. *J. Chem. Theory Comput.* 2011, 7, 4115.
- (37) Hahn, A. M.; Then, H. *Phys. Rev. E* 2009, 80, 031111.
- (38) Shenfeld, D. K.; Xu, H.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *Phys. Rev. E* 2009, 80, 046705.
- 40 (39) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* 2009, 131, 15403.
- (40) Wang, L.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B* 2010, 114, 7294.

REIVINDICACIONES

1. Un método implementado por ordenador de determinación de la intensidad relativa de la unión entre un receptor y miembros individuales de un conjunto de ligandos para formar complejos entre miembros individuales del conjunto de ligandos y el receptor, comprendiendo el método una determinación implementada por ordenador de múltiples diferencias de energía libre de unión relativas para un conjunto de tres o más pares de ligandos que forman al menos un ciclo cerrado, y una determinación implementada por ordenador de la magnitud de la histéresis alrededor de cada ciclo termodinámico cerrado, que comprende:

a) una determinación probabilística implementada por ordenador de las diferencias de energía libre y distribuciones de error alrededor de las diferencias de energía libre a lo largo de cada una de las ramas del ciclo termodinámico cerrado que conducen probabilísticamente al valor de histéresis observado para los ciclos termodinámicos cerrados;

b) una determinación implementada por ordenador de la diferencia de energía libre más probable para cada rama en el ciclo termodinámico cerrado incluido en el modelo probabilístico determinado en la etapa a);

c) una determinación implementada por ordenador del error más probable asociado a la diferencia de energía libre de unión más probable para cada par de ligandos a lo largo de cada rama en el ciclo termodinámico cerrado de la determinación probabilística en la etapa b);

d) una salida implementada por ordenador de valores que representan las energías libres de unión relativas y los errores en la etapa c) a un usuario del ordenador.

2. El método de la reivindicación 1, en el que la etapa de estimación del error comprende el análisis implementado por ordenador de diferencias de energía libre de unión entre ligandos a lo largo de las ramas de más de un ciclo termodinámico cerrado, y la determinación implementada por ordenador de la magnitud de la histéresis en cada uno de los ciclos termodinámicos cerrados.

3. El método de la reivindicación 1, en el que las etapas c y d incluyen la determinación de un conjunto de valores de energía libre para cada rama que minimiza la función

$$\sum_i \frac{(E_i - F_i)^2}{2\sigma_i^2}$$

con la limitación:

$$\sum_i F_i = 0$$

donde E_i es la diferencia de energía libre calculada para una rama i dada;

F_i es la diferencia de energía libre teórica para una rama i dada;

y σ_i es la desviación estándar de las diferencias de energía libre calculadas para la rama i , y donde la suma de las diferencias de energía libre teóricas para todos los ciclos cerrados es 0.

4. El método de la reivindicación 1, donde los ligandos son congéneres.

5. El método de la reivindicación 1, donde se supone una distribución gaussiana en la construcción del modelo probabilístico de la histéresis observada en la etapa a).

6. El método de la reivindicación 1, donde se supone que la distribución de error asociada a las simulaciones de energía libre es uniforme en la etapa a).

7. El método de la reivindicación 1, donde se supone que la distribución de error asociada a las simulaciones de energía libre se suma al error de Bennett en la etapa a).

8. El método de la reivindicación 1, donde la conectividad de los ciclos termodinámicos cerrados se representa como un gráfico.

9. El método de la reivindicación 1, donde la conectividad de los ciclos termodinámicos cerrados se representa como una matriz.

10. El método de la reivindicación 1, donde la determinación probabilística comprende realizar métodos teóricos gráficos.

11. El método de la reivindicación 1, donde la determinación probabilística comprende realizar métodos algebraicos matriciales.
12. El método de la reivindicación 1, donde la determinación probabilística comprende realizar métodos bayesianos.
- 5 13. El método de la reivindicación 1, donde la determinación probabilística comprende realizar métodos de máxima verosimilitud.
14. Un medio de almacenamiento legible por ordenador que comprende instrucciones tangibles no transitorias guardadas en él que son ejecutables por un dispositivo de procesamiento, y con dicha ejecución provocan que el dispositivo de procesamiento realice operaciones según una cualquiera de las reivindicaciones 1 a 13.
- 10 15. Un sistema informático programado con instrucciones no transitorias legibles por ordenador para realizar operaciones según una cualquiera de las reivindicaciones 1 a 13.

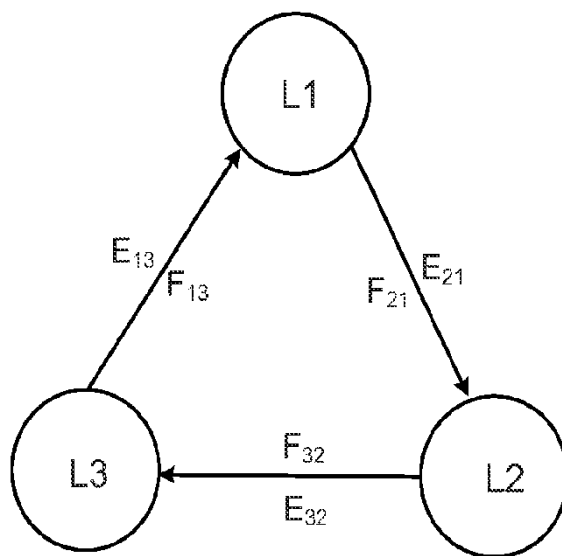


FIG. 1

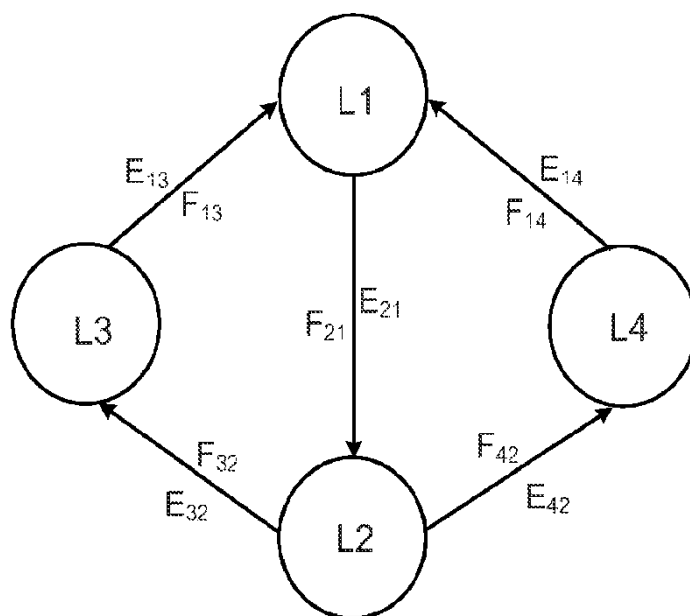


FIG. 2

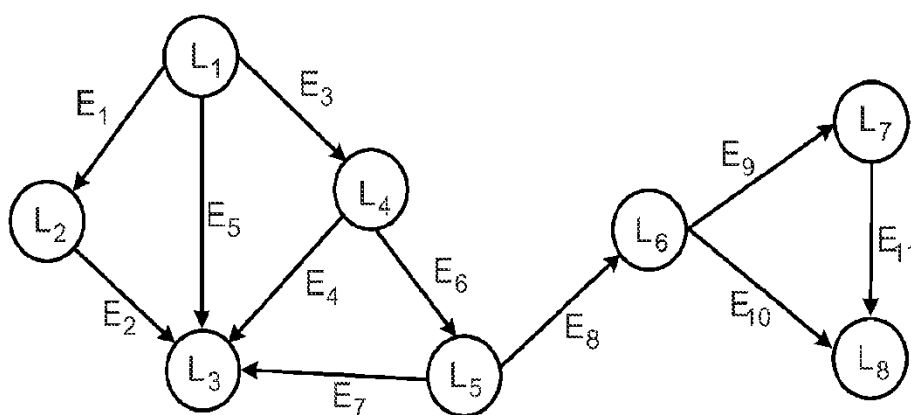


FIG. 3

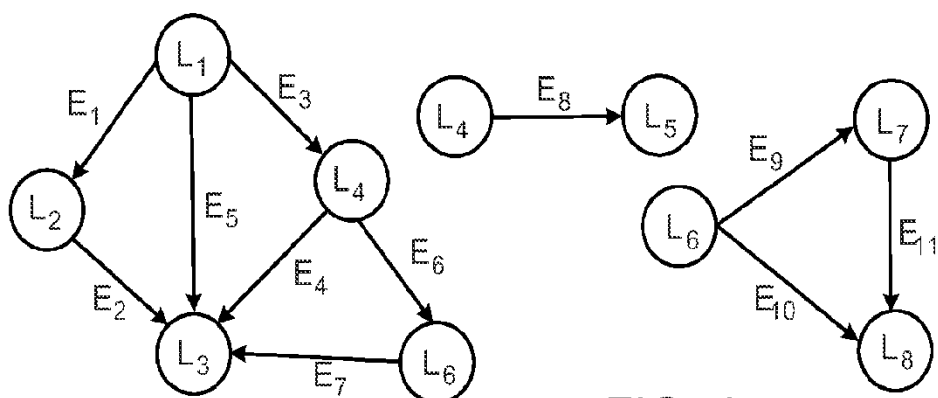
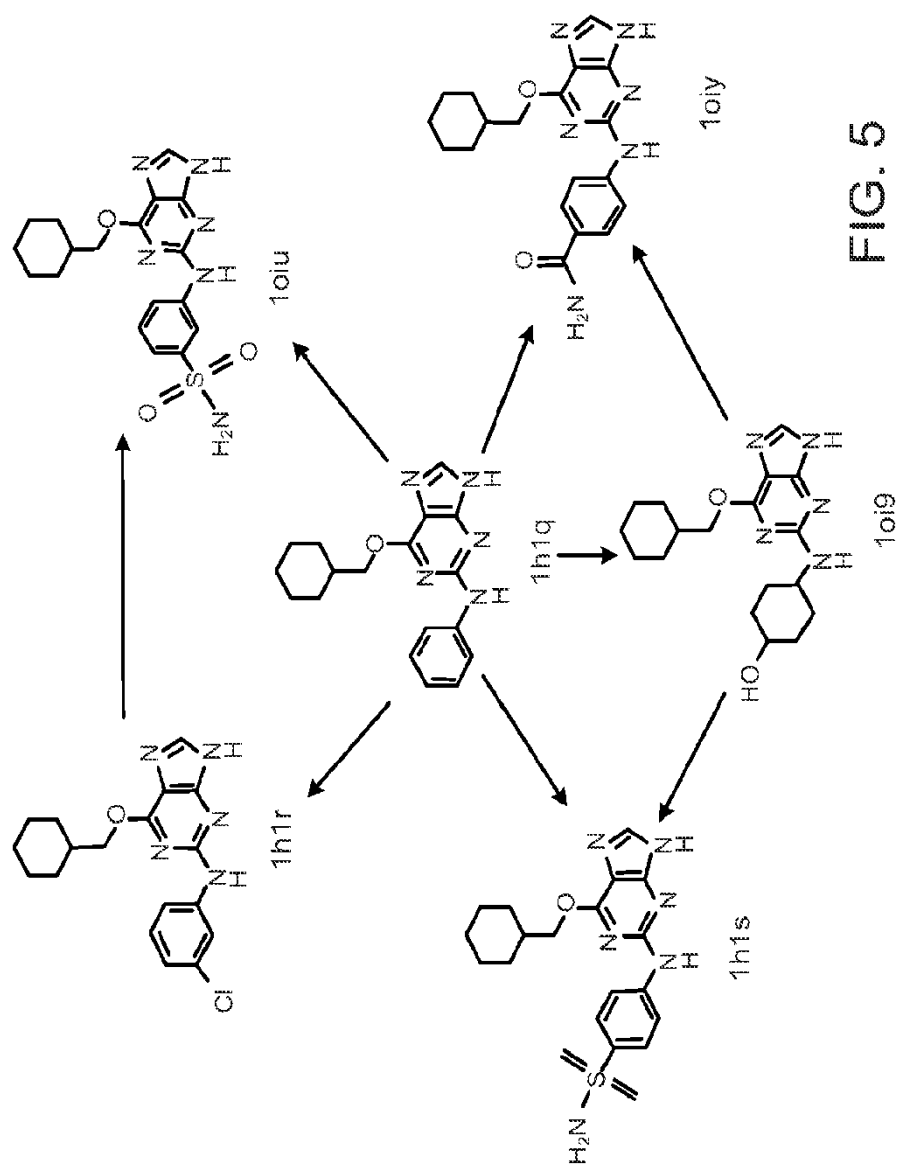


FIG. 4



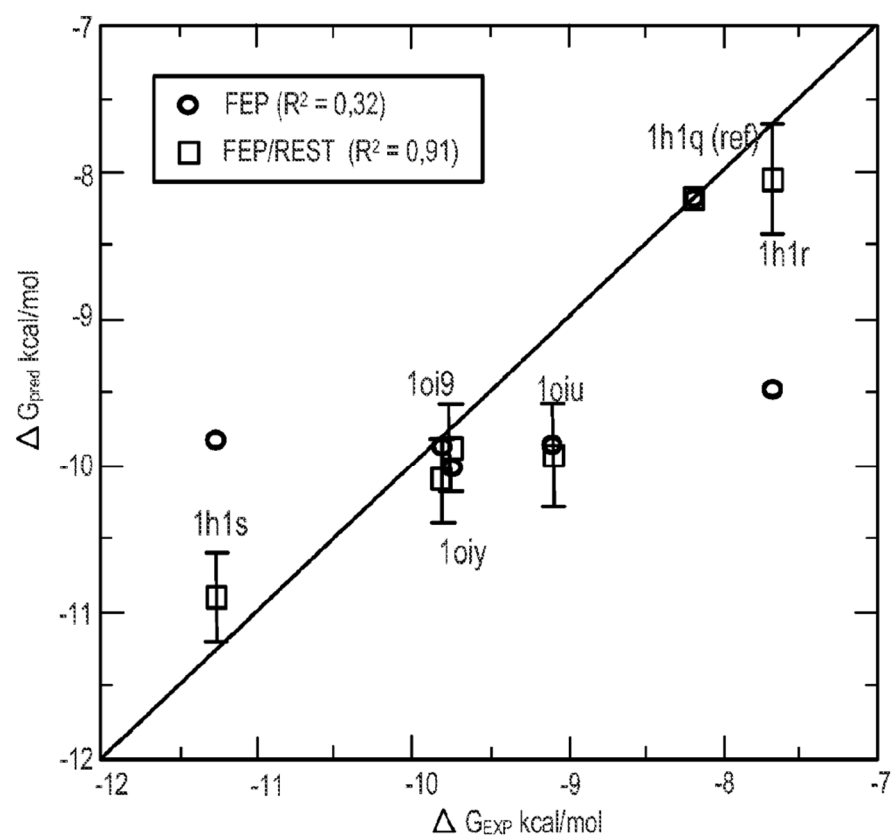


FIG. 6



FIG. 7



FIG. 8

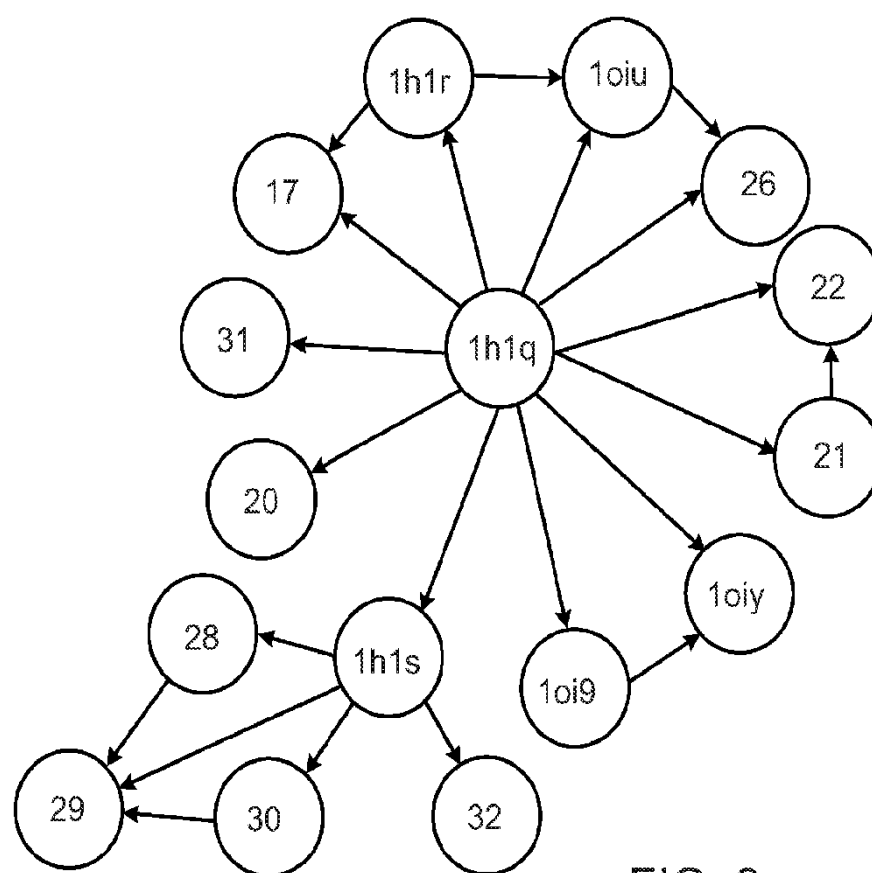


FIG. 9

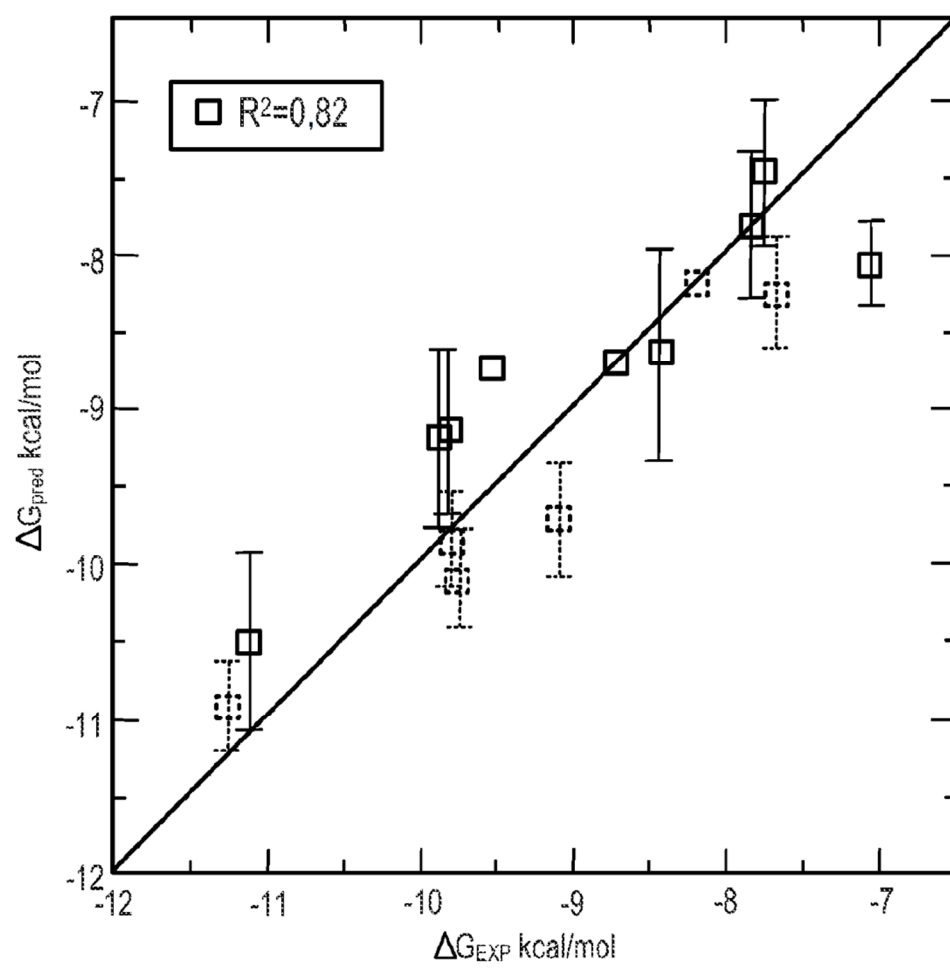


FIG. 10

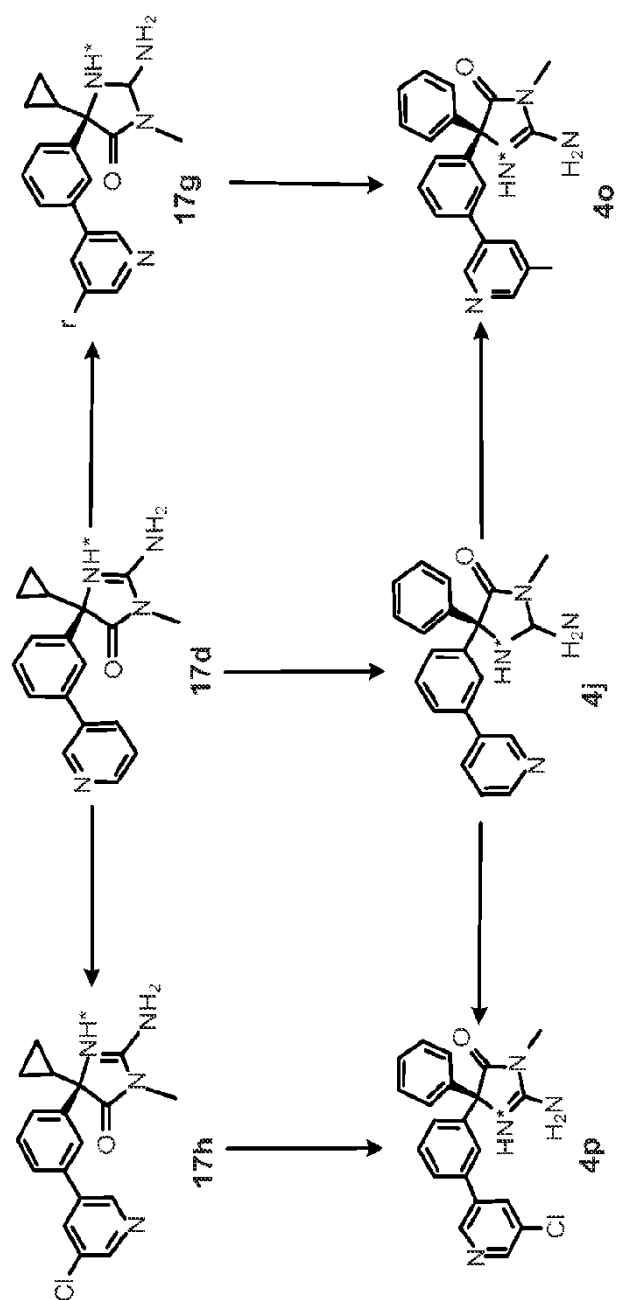


FIG. 11

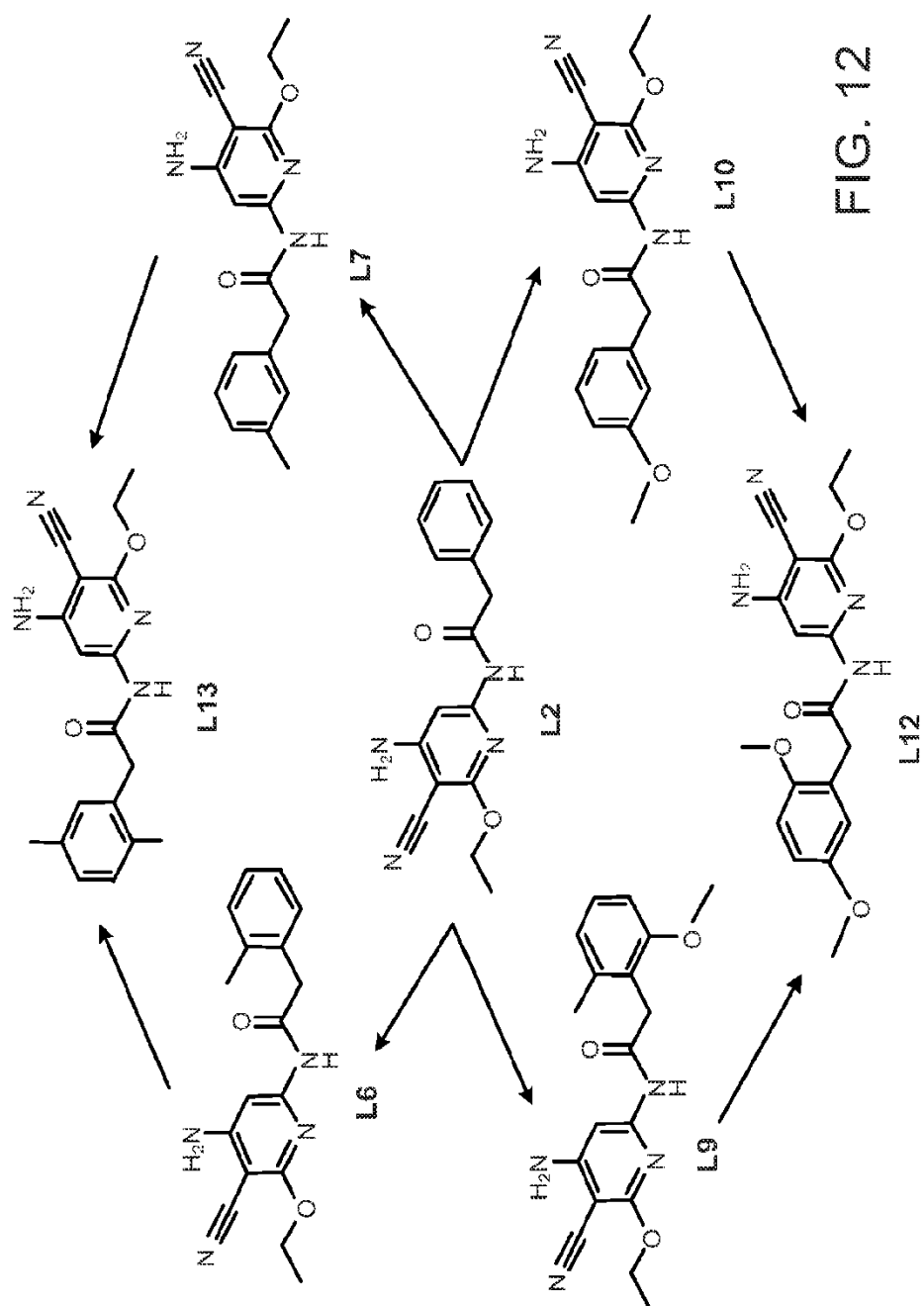


FIG. 12

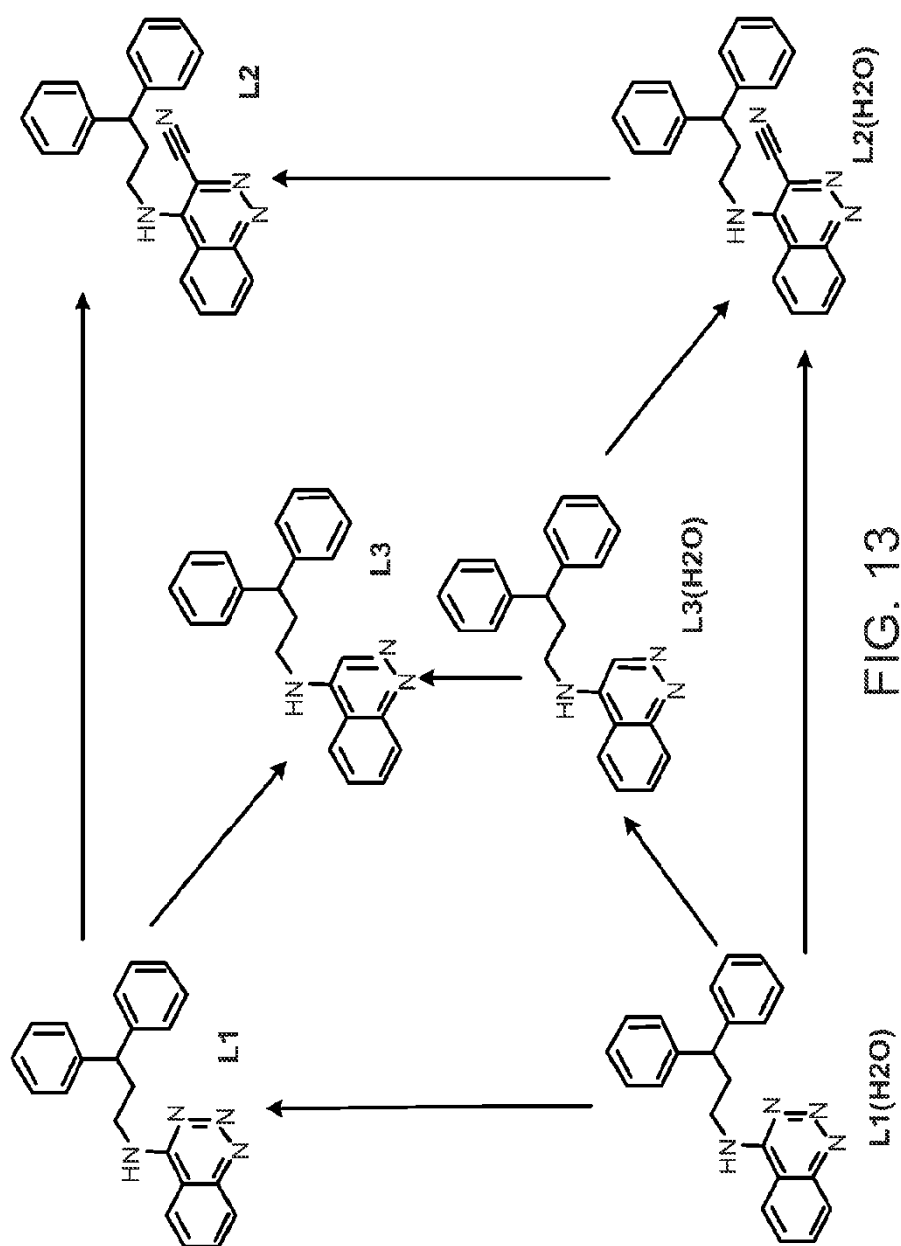


FIG. 13

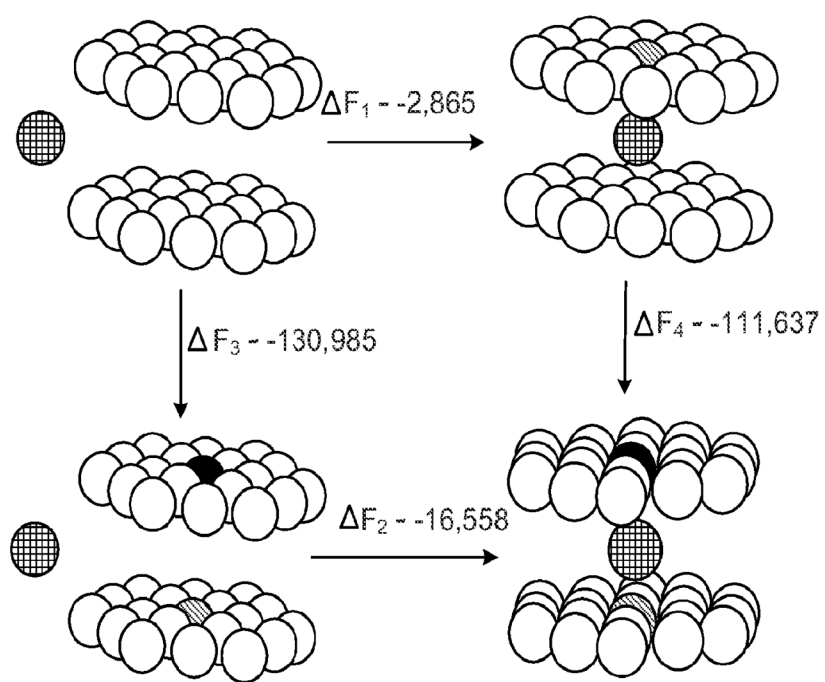


FIG. 14

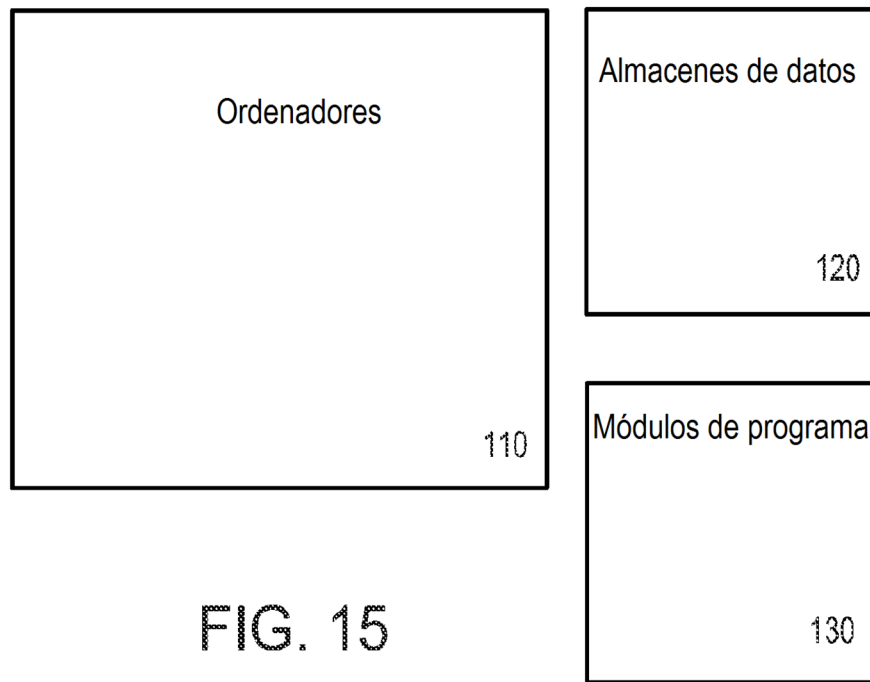


FIG. 15

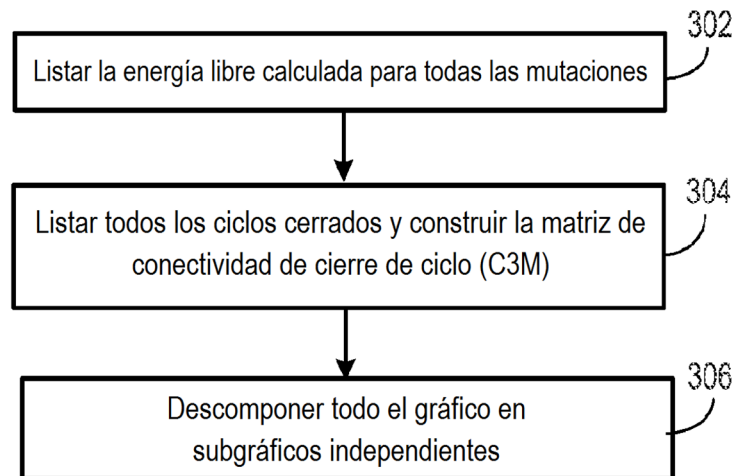


FIG. 16