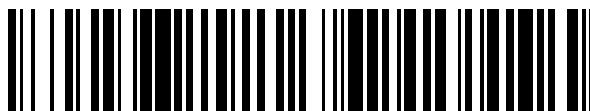


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 432 753**

51 Int. Cl.:

G06F 19/16 (2011.01)

C40B 30/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.03.2006 E 06737802 (6)**

97 Fecha y número de publicación de la concesión europea: **25.09.2013 EP 1856244**

54 Título: **Función de puntuación predictiva para estimar la afinidad de unión**

30 Prioridad:

11.03.2005 US 660822 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.12.2013

73 Titular/es:

**SCHRÖDINGER, LLC (100.0%)
120 WEST FORTY-FIFTH STREET 29TH FLOOR,
TOWER 45
NEW YORK, NY 10036-4041, US**

72 Inventor/es:

**FRIESNER, RICHARD A. y
MURPHY, ROBERT**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 432 753 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

Función de puntuación predictiva para estimar la afinidad de unión

5 Antecedentes de la invención

La invención se refiere a una función de puntuación predictiva para estimar la afinidad de unión.

10 Una gran mayoría de compuestos farmacéuticamente activos se derivan de la eficacia de unión a receptores específicos que a menudo son proteínas (un término que incluye glucoproteínas y lipoproteínas). Hace aproximadamente 50 años, la primera estructura de alta resolución de una proteína se determinó a través de cristalografía de rayos x. El banco de datos de proteínas de acceso público (PDB) contiene actualmente más de 10.000 de dichas estructuras, y las compañías farmacéuticas y de biotecnología tienen un orden de magnitud más de estructuras de propiedad. Muchas de estas estructuras se han cocristalizado con pequeñas moléculas unidas a 15 ellas. El examen de dichas estructuras, y el despliegue del conocimiento adquirido de este modo para diseñar inhibidores nuevos, más potentes, y más específicos, hacen referencia al diseño de fármacos basado en la estructura.

20 El modelado informático facilita el diseño de fármacos basado en la estructura. Dada una estructura de proteína de alta resolución, el software informático se usa para "acoplar" un ligando de molécula pequeña en la posición correcta y orientarlo en la cavidad del sitio activo de la proteína, y calcular una afinidad de unión del ligando dada esta estructura. Programas de software para ordenador que realizan esta tarea se denominan programas de "acoplamiento".

25 Un programa de acoplamiento por lo general realiza dos tareas distintas para modelar la unión proteína-ligando. En primer lugar, se predice una estructura de un complejo de proteína-ligando. En el acoplamiento de alto rendimiento, normalmente se supone que un receptor (por ejemplo, proteína) es rígido. Cuando esta suposición falla, se necesita el uso de una estructura diferente del receptor como un punto de partida. El problema de la construcción de estructuras de receptores alternativos que se modifican para aceptar ligandos que requieren un cambio sustancial en la conformación del receptor ("ajuste inducido") es muy importante. Aquí nos centramos en casos en los que el 30 acoplamiento del ligando en un receptor rígido produce una estructura con una concordancia razonable con los datos experimentales. Este será el caso en el que, en la estructura acoplada, el ligando en la estructura acoplada hace fundamentales los enlaces de hidrógeno y los contactos hidrofóbicos con la proteína en buena concordancia con los datos experimentales.

35 Una segunda tarea del programa de acoplamiento es calcular la afinidad de unión de proteína-ligando, dada como una entrada de la estructura acoplada. Una función matemática usada para calcular la afinidad de unión se denomina "función de puntuación." A pesar del esfuerzo intensivo durante muchas décadas, en muchas situaciones, las funciones de puntuación disponibles actualmente no son satisfactorias con respecto a la precisión y robustez 40 deseadas.

M Stahl y col: "Detailed Analysis of Scoring Functions for Virtual Screening", Journal of Medicinal Chemistry, vol 44, nº 7, páginas 1035-1042 describe funciones de puntuación rápida para el acoplamiento de biblioteca. Se evalúan 45 cuatro funciones de puntuación en siete proteínas diana cuyos sitios de unión representan una amplia gama de tamaño, forma y polaridad, para dar una idea de las fuerzas y debilidades de las funciones de puntuación actuales. R Friesner y col: "Glide: A new Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy", Journal of Medicinal Chemistry, vol 47, nº 7, páginas 1739-1749 describe un método que aproxima una búsqueda sistemática completa del espacio conformacional, orientacional y de posicional del ligando 50 acoplado.

50 Sumario de la invención

La presente invención proporciona un método que se aplica por ordenador para caracterizar la interacción de un 55 ligando propuesto con un receptor especificado de acuerdo con la reivindicación 1. Hay dos aspectos de la divulgación, los cuales presentan un método que se aplica por ordenador para calcular un valor representativo de la interacción (VRI) y un ligando propuesto con un receptor especificado.

Indicado por lo general, en el primer aspecto, las interacciones hidrofóbicas entre uno o más átomos de ligando y 60 uno o más átomos del receptor se puntúan mediante un método que otorga una asignación por la presencia de envoltura hidrofóbica de uno más átomos del ligando por el receptor, y la asignación da como resultado un VRI indicativo de una mayor afinidad de unión ligando/receptor.

En realizaciones preferentes del primer aspecto de la divulgación, el cálculo de VRI incluye al menos un término 65 representativo de energía libre, entalpía o entropía, o una combinación de los mismos. Las interacciones hidrofóbicas de puntuación pueden incluir el cálculo de un componente de puntuación del área superficial (SASC) que varía como una función del área superficial de contacto hidrofóbico-hidrofóbico entre átomos del ligando y

átomos del receptor, y la asignación por la presencia de envoltura hidrofóbica da como resultado un aumento de VRI con respecto a un VRI hipotético que se obtendría usando el SASC pero no la asignación. La puntuación también puede incluir el cálculo de un término de energía del par átomo-átomo, de modo que la asignación por la presencia de envoltura hidrofóbica da como resultado un aumento de VRI con respecto a un VRI hipotético que se obtendría usando la suma del término de energía del par átomo-átomo pero no de la asignación.

Las interacciones hidrofóbicas de puntuación pueden incluir adicionalmente el cálculo de un término de energía del par átomo-átomo, y la asignación por la presencia de envoltura hidrofóbica da como resultado un aumento de VRI con respecto a un VRI hipotético que se obtendría usando la suma del término de energía del par átomo-átomo más el término de SASC, pero no la asignación.

El valor calculado de VRI se puede usar de diversas maneras, por ejemplo, mediante un programa informático en un método para simular el acoplamiento del ligando en el receptor para producir la estructura del complejo receptor-ligando, por ejemplo, para ayudar en la asignación de una estructura al complejo receptor-ligando o para evaluar afinidad de unión del ligando a una estructura del complejo proteína-ligando producida por el programa de acoplamiento.

Como alternativa, el valor VRI se puede calcular en un método que asigna un valor representativo de la interacción receptor-ligando en una estructura predeterminada de receptor-ligando, por ejemplo, una estructura predeterminada de receptor-ligando obtenida a partir de un programa de acoplamiento, a partir de simulaciones dinámicas moleculares o de minimizaciones usando un modelo de solvatación continua.

Preferentemente, una representación gráfica de las características estructurales del complejo receptor-ligando que contribuyen a un mayor VRI debido a la envoltura hidrofóbica se genera cuando la puntuación indica la presencia de envoltura hidrofóbica, y átomos del receptor que contribuyen a la envoltura hidrofóbica de uno o más átomos del ligando se pueden resaltar gráficamente para proporcionar una imagen física de la envoltura. Las superficies del receptor que están implicadas en la envoltura hidrofóbica del ligando también se pueden resaltar en la pantalla.

El primer aspecto de la divulgación se puede usar para seleccionar compuestos para someterlos a ensayo experimentalmente.

Realizaciones preferentes del primer aspecto de la divulgación pueden incluir una o más de las siguientes etapas en cualquier orden (excepto la etapa g) que se realiza al final): a) calcular un término que representa envoltura hidrofóbica; b) calcular un término que representa refuerzos electrostáticos; c) calcular un término que potencia VRI para enlaces de hidrógeno (H) en regiones hidrofóbicas; d) calcular un término que representa pares de enlaces de H; e) calcular términos que representan cationes π y apilamiento π ; f) calcular un término que representa un par lipofílico convencional; y g) calcular la puntuación a partir de una suma de términos. La presencia de envoltura hidrofóbica se puede determinar realizando al menos las siguientes etapas: a) localizar átomos de receptores lipofílicos que están dentro de una distancia especificada del punto de corte de un átomo hidrofóbico del ligando; b) determinar la distribución espacial y angular de átomos del receptor lipofílico alrededor del átomo del ligando hidrofóbico; c) adjudicación de la asignación basándose en la distribución espacial y angular de los átomos del receptor lipofílico alrededor del átomo del ligando hidrofóbico. La presencia y la magnitud de la envoltura hidrofóbica se pueden determinar realizando al menos las siguientes etapas: a) dividir átomos lipofílicos del ligando en grupos de átomos que están conectados por enlaces covalentes; y b) asignar una puntuación de envoltura hidrofóbica a cada grupo, y establecer la puntuación de envoltura a cero si el tamaño del grupo está por debajo de un valor umbral especificado previamente (por ejemplo, mediante la determinación de una puntuación de envoltura hidrofóbica para cada átomo del ligando en el grupo, y a continuación añadir los valores para cada átomo del ligando para obtener la puntuación del grupo); c) calcular la asignación total a adjudicar a la envoltura hidrofóbica como una suma de puntuaciones de envoltura hidrofóbica de los grupos conectados. El valor de la puntuación de la envoltura hidrofóbica para un átomo del ligando en un grupo conectado adecuado se puede determinar realizando al menos las siguientes etapas: a) calcular la distancia entre el átomo del ligando y los átomos lipofílicos del receptor; b) usar una fórmula matemática para evaluar el grado de envoltura del átomo cuando la fórmula tiene en cuenta la dispersión angular de los átomos del receptor lipofílico y su distancia desde el átomo del ligando. Una dispersión angular mayor de los átomos del receptor que están dentro de la distancia del punto de corte especificada del átomo del ligando darán como resultado una mayor puntuación de envoltura usando dicha fórmula, el átomo del ligando que está envuelto por átomos de receptores lipofílicos a distancias adecuadamente cortas producirá una puntuación de envoltura distinta de cero usando dicha fórmula. Además, la puntuación de envoltura hidrofóbica de cada átomo del ligando en el grupo se determina mediante un método que incluye las siguientes etapas: a) designar un átomo de anclaje que es el átomo del receptor lipofílico que está más cerca al átomo de ligando; b) construir un vector de referencia entre el átomo del ligando y el átomo de anclaje; c) construir vectores entre el átomo del ligando y átomo del receptor lipofílicos restantes dentro de una distancia especificada del punto de corte; d) calcular el ángulo formado entre cada uno de los vectores definidos en c) y el vector de referencia; y e) determinar la contribución del átomo del receptor lipofílico que genera el vector en c) a la puntuación de envoltura mediante la distancia del átomo receptor al átomo de ligando, y por el ángulo definido en d), usando una función en la que la contribución aumenta con la disminución de la distancia y con el aumento del ángulo. La contribución a la envoltura para cada átomo del receptor lipofílico es un producto de la puntuación angular y de la puntuación de distancia. La puntuación angular se establece como cero si el ángulo es menor que un ángulo predeterminado, que puede, pero no necesita, variar como

una función de la distancia creciente entre el átomo del ligando y el átomo de anclaje. Átomos de ligando se pueden asignar a un grupo lipofílico que incluye un conjunto de átomos de carbono conectados sin interrumpir mediante conexiones de átomos polares, omitiendo átomos de carbono conectados a átomos polares. Los grupos pueden contener un número mínimo de átomos (por ejemplo, al menos tres átomos). Se puede establecer una condición de que los átomos de receptores lipofílicos que reciben una puntuación de envoltura distinta de cero deben estar dentro de una distancia que es la suma de una constante (por ejemplo, aproximadamente $3 \text{ \AA} \pm 0,1 \text{ \AA}$) más la suma de los radios de van der Waals del ligando y del receptor. La división de átomos en grupos conectados de átomos del ligando pueden implicar una operación matemática que es una función de la conectividad del enlace y una identidad química de cada átomo y sus vecinos.

En una realización preferente, el ligando incluye grupos de átomos neutros y el método incluye: identificar uno o más enlaces de hidrógeno entre los átomos neutros del ligando y el receptor mediante el uso de criterios de distancia y angular; determinar átomos del ligando conectados al grupo o grupos de enlace de hidrógeno; asignar una asignación a enlaces de hidrógeno basándose en (a) si existe envoltura hidrofóbica de los átomos del ligando implicados en los enlaces de hidrógeno, y los átomos del ligando conectados a los grupos de enlace de hidrógeno, y además en (b) si los átomos del ligando identificados en (a) se benefician de la envoltura en la que (i) los átomos del ligando son lipofílicos o (ii) los átomos del ligando han hecho su complemento total de enlaces de hidrógeno sin la envoltura. Se detecta un enlace sencillo de hidrógeno del ligando con el receptor, y se adjudica una asignación al enlace de hidrógeno sencillo especial si ese enlace de hidrógeno y su entorno satisfacen una o más (más preferentemente todas) las siguientes condiciones: (a) un átomo dador o aceptor de enlaces de hidrógeno del ligando es: i) nitrógeno; ii) en un anillo o iii) ambos (i) y (ii); (b) un átomo del ligando forma parte de un grupo de átomos dadores de enlaces de hidrógeno y todos los dadores del grupo están unidos por hidrógeno al receptor; (c) el receptor es una proteína y átomo del receptor que participan en el enlace de hidrógeno son parte de la estructura principal de la proteína; (d) el átomo dador o aceptor del receptor apropiado en la estructura de apoproteína (es decir, la estructura del receptor sin ningún ligando presente) está poco solvatado (por ejemplo, usando un programa de adición de agua basado en la retícula para evaluar la solvatación del átomo dador o aceptor en una estructura de apoproteína del receptor; (e) se calcula la suma de los factores angulares en los términos de envoltura para el átomo pesado del ligando implicado en el enlace de hidrógeno y se calcula cualquiera de los átomos de carbono unidos a este átomo, y la suma es necesaria para superar un umbral especificado (por ejemplo, provocando una pequeña perturbación del ligando y evaluando si la perturbación mejora la puntuación de envoltura para superar el umbral). Por ejemplo, se puede asignar una asignación al enlace de hidrógeno sencillo especial de aproximadamente $-1,5 \pm 0,3 \text{ kcal/mol}$ cuando se satisfacen todos los criterios (a)-(e) anteriores. El criterio (e) se puede imponer tal como se ha descrito anteriormente o por sí mismo. Se pueden necesitar átomos para que tengan menos de tres moléculas de agua en contacto para que cumpla los requisitos para la asignación de enlace de hidrógeno especial. Otros criterios posibles para adjudicar la asignación de enlace de hidrógeno de par especial incluyen uno o más de los siguientes: (a) pares de enlaces de hidrógeno múltiples no cumplen con los requisitos para la asignación de enlace de hidrógeno de par especial cuando los átomos de unión están separados por más de un enlace de rotación, en los que grupos OH cuentan como de no rotación en este cálculo; y/o (b) uno o más de los siguientes tipos de pares de enlaces de hidrógeno del receptor de ligando no cumplen con los requisitos para la asignación de enlace de hidrógeno de par especial:

1) enlaces de H de poca calidad ($< 0,05$ para la puntuación del par HB);
2) pares que implican el mismo átomo de proteína neutra;
3) pares implicados en un puente salino si el potencial electrostático en cada átomo del ligando no está por debajo de un punto de corte;
4) pares del puente salino si el átomo receptor está implicado en un puente salino proteína-proteína;
5) pares dador/dador del ligando que vienen de grupos NH_x $x \geq 2$ en los que N no forma parte de un anillo y tiene carga formal cero;
6) átomos de proteína cargados formalmente con más de 8 moléculas de agua de envoltura secundaria en el estado de no unión;
7) enlaces de H con carga neutra en los que el átomo de proteína no está en un puente salino ;
8) pares de diferentes átomos aceptores neutros en el ligando;
9) Pares de enlaces de H neutros que no cumplen individualmente con los requisitos para un refuerzo de enlace de hidrógeno sencillo, por ejemplo, no satisfacen los controles del átomo del anillo y entorno de fobicidad; y
10) el ligando OH se une al receptor H si el átomo del receptor tiene carga formal cero.

(c) para pares en los que los átomos de ligando del par tienen individualmente carga formal neta cero, a los átomos de hidrógeno se les adjudica una asignación para dicha envoltura hidrofóbica en la que la puntuación de dicha interacción hidrofóbica está optimiza para pares de enlaces de hidrógeno. Las siguientes condiciones se usan en la

etapa de realización (c) anterior para determinar si el par del enlace de hidrógeno tiene una puntuación de envoltura hidrofóbica por encima de una puntuación del punto de corte para recibir la asignación del par de enlace de hidrógeno: a) átomos del ligando en el par deben formar parte del mismo anillo o, para átomos del ligando que no están en el anillo, los átomos del ligando deben estar conectados directamente con el mismo anillo; y b) se detecta hidrofobicidad de una región del enlace de hidrógeno y se añaden factores angulares en términos de envoltura y la suma supera un punto de corte especificado que usa el par de átomos del ligando de unión a hidrógeno y los átomos del ligando en el anillo conectados directamente a los átomos de ligando que participan en el enlace de hidrógeno; y c) si un átomo del ligando del par no es un átomo del anillo pero está conectado con un anillo, la suma anterior incluye átomos del anillo que son vecinos cercanos al átomo del ligando que no está en el anillo. El recuento doble de enlaces de hidrógeno especiales de pares e individuales se puede evitar primero buscando pares y excluyendo cualquier enlace de hidrógeno reforzado encontrado en la lista de pares a partir de la lista de refuerzo del enlace de H sencillo. Criterios adicionales son: d) átomos de ligando en el par deben ser parte del mismo anillo o, para átomos del ligando que no están en el anillo, los átomos del ligando deben estar conectados directamente al mismo anillo; y e) se detecta hidrofobicidad de una región de enlace de hidrógeno y la suma de los factores angulares en términos de envoltura para el átomo pesado del ligando implicado en el enlace de hidrógeno y los átomos del ligando en el anillo conectados directamente con los átomos del ligando que participan en el enlace de hidrógeno, en el que la suma es necesaria para superar un umbral especificado; f) si un átomo del ligando del par no es un átomo del anillo pero está conectado a un anillo, la suma incluye átomos del anillo que son vecinos cercanos al átomo del ligando que no está en el anillo; y g) la asignación adjudicada al par del enlace de hidrógeno es aproximadamente -3 kcal/mol.

Las etapas para evaluar la magnitud de la asignación para la envoltura hidrofóbica son: determinar la presencia de singletes de enlaces de hidrógeno, pares correlacionados de enlaces de hidrógeno y tripletes de enlaces de hidrógeno, y asignar diferentes valores por enlace a singlete, pares, y tripletes de enlaces de hidrógeno. Un valor para enlaces de hidrógeno entre grupos de átomos neutros de un ligando y un receptor se pueden asignar mediante un método que comprende, evaluar el grado de solvatación del receptor, y adjudicar un refuerzo de asignación dependiendo del grado de solvatación del receptor. Dos pares de enlaces de hidrógeno que comparten un átomo en común pueden satisfacer los criterios para la adjudicación de una asignación, y la asignación para el segundo par se puede establecer específicamente, por ejemplo, en aproximadamente -2 kcal/mol. La asignación asignada al par de enlaces de hidrógeno (mencionados en adelante como "refuerzo de base ") se puede reducir de acuerdo con la siguiente fórmula:

- reducir el refuerzo de base usando una función angular:

$$S(\theta) = (\theta - \theta_m) / (\theta_0 - \theta_m), \text{ si } \theta_m < \theta < \theta_0$$

en la que θ_0 es el ángulo máximo y θ_m es el ángulo mínimo permitidos en la parametrización, $S(\theta) = 1,0$, si $\theta \geq \theta_0$ y $S(\theta) = 0,0$, si $\theta < \theta_m$

- asignar a cada enlace del par un factor de escala angular S_1 , S_2 ; y
- aplicar el producto (E_{hb_espec}) del factor de escala angular para cada enlace $S_1 * S_2$ del par al refuerzo de base (E_{base}), de acuerdo con la fórmula $E_{hb_espec} = S_1 * S_2 * E_{base}$.

Por ejemplo, θ_0 se puede establecer en $135^\circ \pm 5^\circ$ y θ_m se puede establecer en $120^\circ \pm 5^\circ$. La solvatación de los compañeros de unión a hidrógeno de la proteína en un enlace de hidrógeno neutro-neutro se puede evaluar aplicando un algoritmo basado en la retícula para añadir moléculas de agua explícitas, y hacer el recuento del número de dichas moléculas de agua en la primera y segunda envoltura alrededor del grupo de proteína en cuestión.

Se puede generar una representación del complejo generada por ordenador en la que enlaces de hidrógeno que reciben una asignación especial debido a la envoltura hidrofóbica están resaltados en una visualización gráfica. La envoltura de enlaces de hidrógeno que reciben una asignación especial debido a la envoltura hidrofóbica también se puede resaltar en la visualización gráfica generada por ordenador, por ejemplo, los átomos del receptor que participan en la envoltura hidrofóbica de enlaces de hidrógeno que reciben una asignación están resaltados en una visualización gráfica. Los átomos del receptor que participan en la envoltura hidrofóbica de enlaces de hidrógeno que reciben una asignación también pueden estar resaltados a través de una visualización de superficies hidrofóbicas del receptor. Los átomos del ligando conectados a enlaces de o enlaces de hidrógeno que reciben una asignación especial debido a la envoltura hidrofóbica también pueden estar resaltados en una visualización gráfica.

Las regiones del receptor capaces de proporcionar envoltura hidrofóbica a grupos de ligando adecuados se pueden detectar realizando simulaciones dinámicas moleculares, e identificando regiones en las que las moléculas de agua tienen potenciales químicos poco favorables.

Las regiones del receptor capaces de proporcionar envoltura hidrofóbica a grupos de ligando adecuados se pueden detectar colocando moléculas de agua rodeando al receptor y detectando entornos envueltos de estas moléculas de agua.

En el segundo aspecto de la divulgación, los enlaces de hidrógeno carga-carga entre un ligando y un receptor están puntuados por: establecimiento de un valor por defecto para un enlace de hidrógeno carga-carga, y adjudicación de una asignación por encima del valor por defecto cuando se satisface uno o más criterios de enlace de hidrógeno carga-carga predeterminado especializado. Los criterios de enlace de hidrógeno carga-carga incluyen uno o más de los siguientes: determinar si un ligando tiene una carga neta o es un zwitterión al determinar si adjudicar la asignación, al determinar el valor de la asignación, o ambos, determinar la distancia entre grupos cargados opuestamente el ligando es un zwitterión, determinar el grado de solvatación del grupo receptor implicado en el enlace de hidrógeno, determinar la geometría detallada del enlace de hidrógeno cargado-cargado, y determinar la energía electrostática del enlace de hidrógeno carga-carga. Por ejemplo, a un zwitterión se le puede asignar una asignación carga-carga adicional si tanto los grupos positivo como negativo del zwitterión forman enlaces de hidrógeno carga-carga, y si la distancia entre los grupos no supera una distancia predeterminada. Las asignaciones pueden aumentar a medida que aumenta la interacción electrostática de los enlaces de hidrógeno carga-carga formados por el zwitterión con el receptor. La asignación también puede aumentar (por ejemplo, de $3,0 \pm 0,5$ kcal/mol a $4,7 \pm 0,5$ kcal/mol) a medida que aumenta la magnitud de las interacciones electrostáticas.

La solvatación del receptor se puede evaluar y se puede asignar un refuerzo de enlace de hidrógeno cargado-cargado si la solvatación del receptor es inferior a un valor predeterminado, por ejemplo, aproximadamente 9 moléculas de agua de envoltura secundarias, tal como se determina mediante un código de adición de agua basado en la retícula, alrededor del átomo del receptor cargado implicado en el enlace de hidrógeno cargado-cargado. Además, se puede asignar un refuerzo de enlace de hidrógeno cargado-cargado adicional si la energía electrostática de receptor y ligando supera un umbral especificado.

Preferentemente, el segundo aspecto de la divulgación incluye determinar si la interacción de grupos cargados positivamente y negativamente en el enlace de hidrógeno cargado-cargado es bidentada o monodentada, y asignar un refuerzo de enlace de hidrógeno cargado-cargado adicional para enlaces bidentados. Por lo tanto, se hace una determinación si grupos con carga múltiple del receptor están implicados en la estructura, y se asigna una asignación especial si grupos con carga múltiple están implicados en la estructura. Si el grupo del ligando es un grupo carboxilato, la asignación asignada para la interacción con grupos del receptor con carga múltiple positiva puede ser $1,0 \pm 0,3$ kcal/mol.

El refuerzo cargado-cargado del zwitterión se determina mediante: determinación del número de moléculas de agua que rodean el componente receptor de los átomos cargados que se unen por puente al receptor y al ligando y eliminación o reducción de refuerzos cargado-cargado en los que las cargas están totalmente hidratadas; y clasificar los enlaces de hidrógeno como: a) monodentado; b) bidentado dentro de un grupo del ligando; o c) bidentado que une por puente de un grupo de ligando a dos grupos de proteínas diferentes y ajustar el refuerzo cargado-cargado basándose en esa clasificación, para reconocer que los enlaces de hidrógeno monodentados tienen un valor inferior que los enlaces de hidrógeno bidentados internos, que a su vez tienen un valor inferior al de los enlaces de hidrógeno bidentados unidos por puente.

Preferentemente, el método incluye el uso de un código de adición de agua basado en la retícula para evaluar la solvatación del compañero del receptor del enlace de hidrógeno cargado-cargado en ausencia de ligando.

Además preferentemente, se da una asignación especial a un átomo cargado en un entorno electrostático que tiene un valor que supera un punto de corte especificado, por ejemplo, $1,5$ kcal/mol \pm $0,3$ kcal/mol.

En otras realizaciones preferentes, un componente de VRI representa enlaces o interacciones de hidrógeno cargado-cargado y ese componente se añade a otros componentes de puntuación para calcular la energía libre total del ligando unido al receptor.

Finalmente, los enlaces de hidrógeno cargado-cargado que reciben la asignación especial pueden estar resaltados en la visualización gráfica.

Al evaluar los enlaces de hidrógeno en el segundo aspecto de la divulgación, puede ser importante evaluar cargas sobre los átomos del ligando. Para cualquier ligando dado, se puede en asignar "cargas particulares" a cada átomo basándose en el campo de fuerza que se está usando para el cálculo del modelado molecular. Un campo de fuerza adecuado es el Potencial Optimizado para el Estado Líquido (OPLS). OPLS y otros campos de fuerza generalmente tienen algoritmos bien definidos para asignar cargas parciales a los átomos.

Las cargas parciales se pueden expresar como alternativa como "incrementos de carga del enlace" y "cargas netas". Los incrementos de carga del enlace son iguales y las cargas opuestas se distribuyen a través de un enlace específico. Los incrementos de carga del enlace, que están compuestos de cargas iguales y opuestas, son neutros por definición. Una vez que los incrementos de carga del enlace se definen (en OPLS, estos incrementos se usan de hecho para construir las cargas parciales), se pueden sumar para dar cargas parciales sobre cada átomo que son las cargas que se obtendrían si no hubiera cargas netas. Las cargas netas entonces son la diferencia entre la carga obtenida al sumar los incrementos de carga del enlace, y las cargas parciales reales.

Un enlace de hidrógeno neutro-neutro es aquél en el que ninguno de los átomos compañeros tiene carga neta definida tal como se ha definido anteriormente. Un enlace de hidrógeno cargado-cargado es uno en el que ambos átomos del ligando y del receptor implicados en el enlace de hidrógeno tienen cargas netas (de signo opuesto).

5

La invención se puede aplicar para realizar una o más de las siguientes ventajas.

Los términos mejorados de una función de puntuación se proporcionan al suponer que la unión en un complejo proteína-ligando está condicionada predominantemente por la búsqueda de moléculas de agua que están mal colocadas en la proteína y su reemplazo con un ligando de una manera adecuada.

10

Los métodos que se describen en el presente documento permiten clasificar semicuantitativamente la capacidad de los ligandos candidatos para unirse a una conformación especificada del receptor de la proteína. El error de la raíz cuadrada del valor medio (RMS) en el cálculo de afinidades de unión de ~120 complejos del PDB de la función de puntuación que usan los nuevos términos es 1,8 kcal/mol, a diferencia de 3,3 kcal/mol para una función de puntuación alternativa cuyo rendimiento es típico para los de la bibliografía. Esta mejora de casi un factor de los en la predicción de la afinidad de unión absoluta se traduce en una capacidad cualitativamente mejor para clasificar compuestos del orden.

15

Los métodos que se describen en el presente documento permiten realizar cálculos virtuales de identificación sistemática en grandes bibliotecas de compuestos (miles de millones de compuestos) y básicamente enriquecer la fracción de compuestos activos en el componente de la puntuación superior de la biblioteca, permitiendo de este modo que los compuestos se seleccionen para evaluación experimental, y evitando la evaluación experimental de compuestos peor clasificados que es mucho menos probable que sean compuestos activos. Los factores de enriquecimiento obtenidos con la nueva función de puntuación a menudo son ~10 veces superiores a los que se han obtenido sin los métodos que se describen en el presente documento.

20

25

Los métodos que se describen en el presente documento permiten realizar la optimización dirigida de un compuesto activo, mediante la predicción de la afinidad de unión a perder o ganar por la sustitución de un grupo funcional del compuesto con una alternativa químicamente accesible. La exactitud mejorada en la predicción de la afinidad de unión obtenida es crítica para liderar los esfuerzos de optimización. Si los errores son demasiado grandes, la probabilidad de tener una molécula mejorada basándose en una función de puntuación se convierte en grande de forma insuficiente para justificar su uso en un contexto práctico.

30

Los métodos que se describen en el presente documento permiten una evaluación de la selectividad de unión de una molécula frente a los dianas de proteína diferentes. La molécula puede estar acoplada en ambas dianas, y las puntuaciones se pueden comparar para estimar si el ligando se une preferentemente a una diana o a la otra.

35

Los métodos que se describen en el presente documento permiten la visualización de características fundamentales del ligando, y del sitio activo de la proteína, y sus interacciones, al proporcionar una interfaz gráfica de usuario (GUI). Las regiones de la envoltura hidrofóbica, y enlaces de hidrógeno neutro-neutro y cargado-cargado especiales, se pueden resaltar usando color, textura, superficies moleculares, y otras características disponibles en la GUI. La visualización de estas regiones y motivos estructurales, que son responsables de mayor afinidad de unión, facilita el diseño de compuestos nuevos, mejorados, ambos con respecto a la mejora de la afinidad de unión, y además con respecto a la retención de la afinidad de unión mientras que se optimizan otras características de la molécula.

40

45

Los métodos que se describen en el presente documento permiten una aproximación coherente, única, en la que algoritmos de muestreo y función de puntuación se optimizan en conjunto.

50

Una aplicación de la invención proporciona todas las ventajas que se ha mencionado anteriormente.

Otras características y ventajas de la invención son evidentes a partir de la siguiente descripción, y a partir de las reivindicaciones.

55 Breve descripción de los dibujos

La Figura 1 es un diagrama de bloques.

60

La Figura 2 es un diagrama de un grupo ligando que interactúa con dos entornos hidrofóbicos distintos.

La Figura 3 es un diagrama de flujo.

La Figura 4 ilustra el ligando BIRB796 activo de Boehringer que se une a la conformación externa de DFG de p38 MAP quinasa (pdb 1kv2).

65

La Figura 5 representa la unión de SB220025 a p38 MAP quinasa.

La Figura 6 ilustra que el inhibidor de Estaurosporina de CDK-2 (pdb 1aq1) se une mediante un par especial de enlace de H rodeado hidrofóticamente a la estructura principal (refuerzo de XP de 3 kc/m) así como el relleno del bolsillo hidrofóbico con el anillo condensado (refuerzo de 4,5 kc/m).

La Figura 7 ilustra cgp6047 inhibidor de CDK-2 acoplado por reticulación en la entrada del pdb de 1aq1.

La Figura 8 ilustra Estreptavidina en biotina (pdb 1stp). Tanto los enlaces de hidrógenos especiales en este complejo, como la envoltura hidrofóbica de la región de unión del hidrógeno, se muestran.

La Figura 9 ilustra el bolsillo HIV-RT (pdb 1rt1) unido con inhibidor mediante enlace de H rodeado hidrofóticamente a la estructura principal (refuerzo de 1,5 kc/m) y empaquetamiento hidrofóbico del anillo aromático del ligando (refuerzo de 4,5 kc/m).

La Figura 10 ilustra un inhibidor de Trombina (pdb 1ett) unido mediante un puente salino a ASP_189 de rebaje.

La Figura 11 ilustra un ligando unido a la proteína neuramididasa (pdb 1bji).

Al igual que los números de referencia y designaciones, en los diversos dibujos se indican elementos similares.

Descripción detallada

Tal como se muestra en la Figura 1, un diagrama de bloque de una geometría de interacción molecular a modo de ejemplo incluye una primera entidad molecular 20 y una segunda entidad molecular 22. En algunos casos, la primera entidad molecular 20 puede ser una molécula de fármaco candidata (es decir, ligando), y la segunda entidad molecular 22 puede ser una proteína. La primera entidad molecular 20 incluye los átomos 24, 28, 30 y 32, con sus cargas electrostáticas respectivas (o cargas parciales) de positiva, positiva, positiva, y negativa. La segunda entidad molecular 22 incluye los átomos 34, 36 y 38, con sus cargas electrostáticas respectivas (o cargas parciales) de positiva, negativa, y negativa. Se considera que la entidad 20 está acoplada con la entidad 22. Una afinidad de unión proteína 22 - ligando 20, dada como una aportación de la estructura acoplada que se muestra en la Figura 1, se puede calcular usando una función de puntuación. Una función de puntuación proporciona una forma de calcular la afinidad de unión de un ligando receptor de proteínas, proporcionando una estructura lo suficientemente precisa del complejo. Las afinidades calculadas también se pueden usar para clasificar ligandos de orden.

El uso eficaz de una función de puntuación de alta precisión requiere por lo general una metodología de muestreo mejorada o las estructuras acopladas resultantes son incapaces de evitar grandes penalizaciones por desolvatación, o localizar los motivos estructurales específicos que se han identificado como que contribuyen a una mayor afinidad de unión. Aquí, suponemos que las estructuras de esta calidad se generan para aplicar de forma eficaz la función de puntuación.

Nuestra función de puntuación clasificar semicuantitativamente una capacidad de ligandos candidatos para unirse a una conformación especificada de un receptor de proteína. No se puede esperar que los ligandos que no "se ajustan" en el receptor, debido a choques estéricos significativos, consigan buenas puntuaciones, incluso si se unen eficazmente a una conformación alternativa del mismo receptor.

Las definiciones de "conformación" del receptor y "ajuste" incluyen un componente heurístico. Se tiene que permitir un cierto nivel de desviación de una conformación real inducida por el ligando desde el receptor rígido que se está bloqueando mediante una función de energía potencial usada para acoplamiento; de otro modo, un modelo no sería útil para identificación sistemática virtual. El algoritmo de acoplamiento usado para generar estructuras del complejo proteína-ligando requiere aumentar los radios de van der Waals de los átomos de la proteína y/o ligando, construyendo modestos efectos "de ajuste inducido". Existen muchos casos en los que un grado razonable de aumento no permitirá que el ligando se acople correctamente (por ejemplo, si una cadena lateral está en un estado rotámero que es muy diferente al del complejo nativo proteína-ligando y bloquea a los átomos del ligando frente a la ocupación de su ubicación preferente en un bolsillo de unión). Siempre habrá situaciones en el límite, pero en la práctica hemos encontrado que es posible asignar una gran mayoría de los casos de acoplamiento cruzado como "ajuste" o "no ajuste."

La descripción en el presente documento se centra en complejos en los que en los que el ligando se ajusta correctamente en el receptor, tal como se juzga mediante una capacidad para hacer fundamental la formación de enlaces de hidrógeno y contactos hidrofóbicos, y conseguir una distancia de la raíz cuadrada del valor medio (RMSD) razonable, en comparación con el complejo nativo (o tal como se obtiene por analogía con el complejo nativo o un ligando relacionado).

Factores potenciales a afinidad de unión a proteína-ligando incluyen los siguientes cinco factores:

En primer lugar, un término que representa el desplazamiento de moléculas de agua por el ligando desde

"regiones hidrofóbicas" del sitio activo de la proteína es un factor potencial. La liberación de estas moléculas de agua en disolución puede dar como resultado una disminución de la energía libre global. Las moléculas de agua en dichas regiones pueden no ser capaces de formar su complemento total de enlaces de hidrógeno que estarían disponibles en disolución. También puede haber consideraciones entrópicas. Si una molécula de agua tiene una movilidad restringida en una cavidad de proteína, la liberación en el disolvente dará como resultado una ganancia de entropía. Dado que un ligando libera muchas moléculas de agua, este término contribuirá favorablemente a la energía libre.

En segundo lugar, un término que representa la formación de enlace de hidrógeno entre el ligando y la proteína es un factor potencial. Si los enlaces de hidrógeno formados por el ligando y la proteína son "mejores" en algún modo que los formados entre cada resto y agua, dará como resultado ganancia de energía libre. Una razón por la que dicha sustitución puede ser favorable es si la unión de agua con la proteína o ligando está mal posicionada de otro modo para formar enlaces de hidrógeno adicionales. Este término tiene cierta superposición con el término que representa el desplazamiento de las moléculas de agua por el ligando desde las "regiones hidrofóbicas" del sitio activo de la proteína. Sin embargo, las interacciones por formación de enlace de hidrógeno se tienen que tratar específicamente, al menos en una función de puntuación empírica del tipo que consideramos aquí.

En tercer lugar, un término que representa efectos de desolvatación puede ser un factor o factores potenciales. Los grupos en el ligando o proteína, expuestos anteriormente al agua, pueden quedar sin solvatar al ser colocados cerca de grupos con los que no pueden formar enlace de hidrógeno. Dichos efectos sólo pueden servir para reducir la afinidad de unión, en contraste con los dos términos que se han descrito anteriormente.

En cuarto lugar, un término que representa efectos entrópicos debido a la restricción después de la unión por el movimiento de la proteína flexible o grupos de ligando puede ser un factor potencial. Tal como en el caso de términos de desolvatación, dichos efectos servirán exclusivamente para reducir la afinidad de unión.

En quinto lugar, se necesitan términos especializados para describir la interacción del ligando con iones metálicos. Este término se puede añadir a otros términos en la función de puntuación si un átomo metálico está presente en la proteína. Aquí no consideramos los átomos metálicos. Las realizaciones de la invención se aplican a proteínas que contienen metales, pero la función de puntuación completa, final se debe complementar con un término especializado para metales tal como se ha descrito anteriormente.

Un gran número de funciones empíricas de puntuación se han presentado anteriormente. Mientras que se diferencian unas de otras en detalle, estas funciones de puntuación anteriores tienen similitudes entre sí. Una de dichas funciones de puntuación anteriores que usamos como un ejemplo para la comparación de nuestra función de puntuación es ChemScore. La función de puntuación ChemScore incluye interacciones entre átomos lipofílicos, dadores y aceptores de enlace de hidrógeno, y aceptores y metales de enlace de hidrógeno, y una contribución adicional basándose en el número de enlaces de rotación restringidos del ligando. Cuando se usa ChemScore, la posición de los hidrógenos implicados en la formación de enlaces de hidrógeno se optimiza con respecto a la energía del enlace de hidrógeno (esto se hace para los dadores tanto de proteína como de ligando).

ChemScore trata cada uno de los cuartos términos que se han presentado anteriormente (excluyendo el término específico de átomos metálicos) tal como sigue a continuación.

ChemScore contiene un término de energía del par átomo-átomo hidrofóbico E_{ph} de la forma:

$$E_{ph} = \sum_{ij} f(r_{ij})$$

en la que 'i' y 'j' se refieren a átomos de carbono (lipofílicos), y ' $f(r_{ij})$ ' es una función lineal de la distancia interatómica ' r_{ij} '. Para r_{ij} menores que la suma de los radios atómicos de Van der Waals (Vdw) más 0,5 angstrom, f es 1,0. Para distancias mayores f aumenta linealmente de cero a 3 Å más allá.

Este término E_{ph} representa heurísticamente el desplazamiento de las moléculas de agua de las regiones hidrofóbicas, lo que se consigue más fácilmente mediante átomos de ligandos lipofílicos. Numerosos contactos fuertes entre ligandos lipofílicos y átomos de proteínas indican que las moléculas de agua poco solvatadas se desplazan satisfactoriamente por átomos lipofílicos del ligando, que en sí mismos se expusieron anteriormente al agua. La segregación resultante de átomos lipofílicos, y la liberación simultánea de moléculas de agua de este sitio activo a la mayor parte, disminuye la energía libre a través del efecto hidrofóbico, que se captura aproximadamente mediante la función de puntuación anterior. Los términos que implican contacto entre superficies hidrofóbicas de la proteína y ligando, mientras que difieren en los detalles, intentan básicamente medir el mismo cambio de energía libre a través de un tipo de aproximación muy similar.

ChemScore evalúa la calidad del enlace de hidrógeno proteína-ligando basándose en criterios geométricos, pero no distingue entre diferentes tipos de enlaces de hidrógeno, o entornos de proteínas diferentes.

ChemScore no trata efectos de desolvatación.

5

ChemScore usa un término sencillo de enlace giratorio para tratar efectos de entropía conformacional en el ligando.

Nuestra función de puntuación comienza a partir de los términos usados en ChemScore analizados anteriormente, aunque su parametrización detallada específica para nuestra función de puntuación. Nuestra función de puntuación revisa cualitativamente el término hidrofóbico y los términos de formación de enlace de hidrógeno. También hemos desarrollado términos de desolvatación. Los términos de desolvatación se deben incluir para calcular la afinidad de unión total.

10

Nuestra función de puntuación incluye un modelo mejorado de interacciones hidrofóbicas a las que hacemos referencia como "envoltura hidrofóbica." La función del par átomo-átomo "convencional" tal como se ilustra en ChemScore, E_{PH} , a partir de la que empezamos, asigna una puntuación a un ligando lipofílicos basándose en la distribución de la distancia de átomos vecinos de proteínas lipofílicas. Esto captura claramente un componente significativo de la física del componente hidrofóbico de la unión del ligando. El desplazamiento de las moléculas de agua de áreas con muchos átomos de proteína lipofílica cercanos probablemente tiene más valor en energía libre que en desplazamiento de áreas con menor número de dichos átomos. Como un ejemplo, es evidente que si un ligando esta colocado en una cavidad del sitio activo, en oposición a la superficie de la proteína, los átomos lipofílicos del ligando son propensos a recibir mejores puntuaciones. Si están ubicados en un "bolsillo hidrofóbico" de la proteína, las puntuaciones serían mejores que en una ubicación rodeado principalmente por grupos polares o cargados. Sin embargo, una función de puntuación dependiente solamente de la distribución de la distancia no es sensible a los detalles de la geometría local de los átomos de proteínas con respecto al átomo lipofílico del ligando en cuestión.

15

20

25

Como un ejemplo, considerar dos distribuciones del modelo que se muestran en la Figura 2. En un caso (A), un átomo del ligando 50 esta colocado en una "pared" hidrofóbica. Los átomos de proteína lipofílica 52 solo están en un "lado" del átomo hidrofóbico. En el segundo caso (B), el átomos del ligando lipofílico 54 está ubicado en un bolsillo estrecho, con átomos de proteína lipofílica 56 en dos (o tres) "lados" del átomo del ligando 56. Ajustando los parámetros de distancia y los números de los átomos de proteína apropiados, es posible construir dos modelos que proporcionan fluctuaciones idénticas para cada término del par átomo-átomo (independientemente de los coeficientes) para las estructuras A y B. La cuestión es si la contribución del átomo a la energía libre de unión en las construcciones A y B sería idéntica en las condiciones estipuladas.

30

35

La respuesta depende principalmente de la energía libre que se puede obtener mediante el desplazamiento de una molécula de agua en una ubicación dada en una cavidad de proteína, que depende del grado de satisfacción con el que el agua es capaz de satisfacer sus requisitos de formación de enlace de hidrógeno en esa ubicación. En un caso extremo, en el que una sola molécula de agua se coloca en una cavidad de la proteína que puede alojar solamente agua, y está rodeada por todos los lados con átomos lipofílicos que no pueden formar enlaces de hidrógeno, la energía libre de transferencia al volumen a través del desplazamiento por un grupo de ligando hidrofóbico es claramente favorable, y no está claro que una molécula de agua ocuparía dicha cavidad con preferencia a dejar un vacío, a pesar de los términos estadísticos que favorecen la ocupación. Éstos una situación muy extraña que no es particularmente apropiada para formar enlaces con un ligando grande, hinchas que los ejemplos en la Figura 2 son bastante habituales.

40

45

Hemos realizado un gran número de experimentos computacionales que implican modificaciones del término de puntuación hidrofóbico (en nuestra función de puntuación) diseñados para discriminar entre diferentes entornos de proteínas geométricas. Un criterio para el éxito en estos experimentos es la capacidad de cualquier término de puntuación hidrofóbico propuesto para ajustarse a un amplio intervalo de datos experimentales de energía libre sin unión y proporcionar buenas predicciones para estudios de enriquecimiento. Nuestros resultados se describen a continuación en los párrafos siguientes.

50

Los átomos lipofílicos del ligando se deben considerar en grupos, en lugar de individualmente. La energía libre de las culturas de agua en una cavidad de la proteína se ven afectadas adversamente más allá de la norma principalmente cuando se colocan en un microentorno hidrofóbico extendido sobre la dimensión de varios átomos. Si hay contactos hidrofóbicos aislados individuales, el agua por lo general será capaz de formar su complemento de enlaces de hidrógeno de cualquier modo, evitando estos contactos y usando moléculas de agua vecinales como asociadas. Hemos establecido en tres un tamaño de grupo mínimo de átomos del ligando lipofílico conectados.

55

60

Cuando un grupo de átomos del ligando lipofílico está incluido en dos lados (con un ángulo de 180 grados) por átomos proteína lipofílica, nuestro método hace la hipótesis de que este tipo de estructura contribuye a la energía libre de unión más allá de lo que está codificado en el término del par átomo-átomo. Hacemos referencia a esta situación como "envoltura hidrofóbica" del ligando. El término del par átomo-átomo se ajusta a un conjunto de datos heterogéneo, enorme y está dominado por átomos de ligando lipofílico individuales en un entorno "promedio", es

65

decir, no en un grupo que reside en una región hidrofóbica configurada específicamente de la proteína. Estos términos de la función de puntuación están diseñados para capturar desviaciones favorables desde este promedio, que representan motivos específicos de reconocimiento molecular. Es decir, la colocación de un grupo del ligando hidrofóbico apropiados dentro de una región de la proteína especificada conduce a aumentos sustanciales en la potencia más allá del promedio para dicho grupo, del tipo que por lo general son objetivos de programas de optimización en química medicinal.

Tal como se muestra en la Figura 3, un proceso de puntuación de empaquetamiento hidrofóbico 100 incluye dividir (102) átomos lipofílicos en grupos conectados.

Para cada átomo en un grupo conectado, el proceso 100 enumera (104) átomos de proteínas lipofílicas a diversas distancias. El proceso 100 selecciona (106) el átomo de proteína lipofílica más cercano, y dibuja (108) un vector entre él y el átomo del ligando. Este se denomina átomo de "anclaje".

Todos los otros átomos en la lista se toman, uno a la vez, y para cada átomo, el proceso 100 dibuja (110) un vector para el átomo de ligando, y calcula (112) un ángulo entre los vectores a los átomos de ligando y de anclaje, usando un simple punto del algoritmo del producto. Con el fin de ser considerado en el "otro lado" del átomo de anclaje, el ángulo entre el vector debe exceder algún valor (que depende de la distancia - cuanto más lejano, el ángulo tiene que estar más cerca a 180 grados). Si el ángulo es 0 grados, entonces los vectores son paralelos, y el átomo está en el "mismo sitio." Si el ángulo es de 90 grados, los vectores están en ángulos rectos y el átomo está "en ángulos rectos" con el anclaje. Si el ángulo es de 180 grados, el átomo está exactamente en el "otro lado" y éstos lo que conduce a un entorno especialmente malo para las moléculas de agua. Esto está apoyado por simulaciones dinámicas moleculares de moléculas de agua en superficies hidrofóbicas de curvatura variable, y también por los argumentos geométricos del sentido común.

El proceso 100 asigna a cada átomo una puntuación basándose en el número de contactos lipofílicos totales (con un punto de corte de distancia), denominado puntuación de contacto, y una ponderación que depende del término del ángulo. Si, por ejemplo, ningún átomo está a más de 90 grados del anclaje, el término del ángulo es cero y el átomo contribuye al término "hidrofóbico especial" del grupo. El proceso 100 añade puntuaciones de todos los átomos en el grupo, y el total es la puntuación para ese grupo. Una descripción matemática detallada del proceso 100 se proporciona en el Apéndice I.

Si la puntuación es superior a 4,5 kcal/mol, el proceso 100 limita el término a este valor. Esta es una determinación empírica basándose en el análisis de muchos casos diferentes de ensayo y a la comparación de los resultados con datos experimentales. Si una región de este tipo es muy grande, lo que lleva a una puntuación superior a 4,5, probablemente existe alguna capacidad de las moléculas de agua para compensar mediante la interacción entre sí - regiones más pequeñas tienen más dificultades hacer ésto, el término de limitación compensa este efecto.

Una ilustración experimental en particular de la ganancia de la afinidad de unión si un grupo hidrofóbico grande (anillos en ambos casos) está colocado en un bolsillo en el que átomos de proteínas lipofílicas están presentes en ambos lados del bolsillo se muestra en la Figura 4. Se mostró que el grupo naftilo 60 de este enlazador muy ajustado mejora la unión con respecto a un grupo fenilo mediante un factor de 21. El grupo naftilo 60 se necesita para ocupar totalmente el bolsillo fóbico 62 representado en la Figura 4.

Tal como se ha descrito anteriormente, hacemos referencia a los alrededores de átomos o grupos lipofílicos del ligando mediante átomos de proteínas lipofílicas como una "envoltura hidrofóbica." El tratamiento apropiado de la envoltura hidrofóbica es fundamental para la discriminación de motivos y compuestos de unión altamente y débilmente potentes. La optimización detallada de los criterios numéricos para reconocer envolturas hidrofóbicas, y la asignación de una contribución específica a la afinidad de unión para cada motivo, es importante si se desea desarrollar métodos con capacidad predictiva, y hemos afinado los términos de nuestra función de puntuación usando un gran conjunto de datos experimentales.

Lo anterior describe un método en particular para detectar envolturas hidrofóbicas, partiendo de una estructura predicha de un complejo proteína-ligando. Se puede construir otros métodos para conseguir casi los mismos resultados con respecto al cálculo de la contribución de la envoltura a la afinidad de unión.

El método que se ha descrito anteriormente para la detección de envolturas hidrofóbicas se basa en el uso del complejo proteína-ligando acoplado como un punto de partida para el cálculo de la envoltura. Otro enfoque es comenzar con una estructura de apoproteína, es decir, la proteína en ausencia de cualquier ligando, y realizar una simulación de la proteína usando moléculas de agua explícitas como disolvente. Se puede detectar si existía cualquier ubicación en el sistema en el que las moléculas de agua tenían dificultades para formar su complemento total de enlaces de hidrógeno debido a la presencia de envolturas hidrofóbicas por la proteína, y marcar estas regiones como favorables para la ocupación con grupos de ligandos apropiados. Este enfoque usa características de la formación de enlaces de hidrógeno de moléculas de agua en ubicaciones específicas para detectar la envoltura, en oposición a la inferencia de dichas características a partir del complejo proteína-ligando acoplado. Basándose en dichas simulaciones, se puede conocer términos de una función de puntuación que refuercen la ocupación de las

regiones de proteínas adecuadas (por ejemplo, en un esquema basado en la retícula).

Nuestra función de puntuación incluye un término que representa un modelo mejorado de formación de enlaces de hidrógeno de proteína-ligando. En el desarrollo de un modelo refinado de formación de enlaces de hidrógeno, dividimos los enlaces de hidrógeno en tres tipos: neutro-neutro; neutro-cargado; y cargado-cargado. El análisis de cada tipo de enlace de hidrógeno es diferente, debido a problemas asociados con el intervalo largo de energía desolvatación (Energía de Born) de grupos cargados. Una etapa inicial es asignar diferentes valores por defecto (suponiendo características geométricas óptimas) para cada uno de estos diferentes tipos de enlaces de hidrógeno. Los valores por defecto que asignamos son: neutro-neutro, 1 kcal/mol; neutro-cargado, 0,5 kcal/mol; y cargado-cargado 0,0 kcal/mol. Estas asignaciones se basan en una combinación de observaciones empíricas (mejor ajuste a afinidades de unión indicadas de una amplia gama de complejos del Banco de Datos de Proteínas (PDB)) y razonamiento físico.

Una razón para reforzar enlaces de hidrógeno proteína-ligando en lo más mínimo es útil porque cualquiera de dichos enlaces de hidrógeno está reemplazando enlaces de hidrógeno que la proteína y ligando están formando con agua. El número neto de enlaces totales de hidrógeno como promedio permanecerá igual en el complejo unido en comparación con la solución. Sin embargo, la liberación de moléculas de agua al volumen da como resultado un aumento de entropía, y la liberación de moléculas de agua alrededor de un grupo polar requiere un enlace de hidrógeno proteína-ligando. Este análisis es más verosímil cuando ambos grupos son neutros. La formación de un puente salino entre proteína ligando implica tipos muy diferentes de enlaces de hidrógeno de lo que está en solución. La termodinámica de formación de puentes salinos se ha estudiado proteínas, y depende de muchos factores tales como el grado de exposición al disolvente. El valor por defecto de cero que asignamos se basa en la presencia de muchos complejos proteína-ligando en el PDB con afinidades de enlace muy bajas en los que se forman puentes de hidrógeno. Se necesitan características especiales de un puente salino para que este tipo de estructura contribuya a la afinidad de unión en nuestra función de puntuación. Finalmente, el valor por defecto cargado-neutro representa una interpolación entre el valor neutro-neutro y cargado-cargado, que parece ser consistente con los datos empíricos.

Las puntuaciones de los enlaces de hidrógeno disminuyen de sus valores por defecto a medida que la geometría se desvía de una geometría ideal del enlace de hidrógeno, basándose en ambos ángulos entre los átomos dador y aceptor y la distancia. Una descripción detallada se proporciona en el Apéndice II. A continuación describimos estructuras especializadas en las que a los enlaces de hidrógeno se les asigna incrementos adicionales de afinidad de unión (denominados "refuerzos") que están significativamente en exceso sobre los valores por defecto. Estas situaciones pueden surgir para los enlaces de hidrógeno neutro-neutro, o cargado-cargado, pero no para los enlaces de hidrógeno cargado-neutro. La exclusión de refuerzos especiales de enlace de hidrógeno cargado-neutro ha sido impulsada por el fallo al identificar motivos de este tipo que puedan producir mejor concordancia con los datos experimentales.

Ahora describimos tipos particulares de enlaces de hidrógeno neutro-neutro que se han identificado, basándose en consideraciones tanto teóricas como empíricas, como contribuciones excepcionales a la afinidad de unión. Dichos enlaces de hidrógeno "especiales" representan motivos fundamentales de reconocimiento molecular, y se encuentran en muchas, sino en la mayoría de las dianas farmacéuticas. La resolución de dichos motivos es una estrategia central en el aumento de la potencia y la especificidad de compuestos medicinales. El desarrollo de una función de puntuación automatizada capaz de asignar correctamente la afinidad de unión a estructuras de este tipo (como oposición posterior al hecho de la racionalización de enlaces de hidrógeno "importantes" en cada diana individual) es un objetivo importante en el desarrollo de funciones de puntuación empíricas funciones de puntuación, que permite una mejora espectacular en la capacidad de predicción tanto cualitativa como cuantitativa.

Una idea fundamental con respecto a los enlaces de hidrógeno importantes es localizar posiciones en la cavidad de la proteína en la que las moléculas de agua que forman un enlace de hidrógeno con un grupo polar tendrían una dificultad particular al formar su complemento de enlaces de hidrógeno adicionales. La formación de un enlace de hidrógeno con un átomo de proteína impone restricciones geométricas no triviales sobre la molécula de agua. Esta es la base para la puntuación del enlace de hidrógeno por defecto. Pero dichas restricciones se hacen más problemáticas cuando el entorno restante de la molécula de agua se estimula con respecto a formar enlaces de hidrógeno.

Nuestro análisis de interacciones hidrofóbicas sugiere que el entorno se hará cualitativamente más exigentes y la molécula de agua tiene átomos de proteínas hidrofóbicas en los dos lados (en lugar de solo uno) y también si hay pocas moléculas de agua cercanas con la capacidad de reajustarse a sí mismas a la geometría restringida del enlace de hidrógeno proteína-ligando. Identificamos geometrías de este tipo mediante el uso de una variante del método que se ha descrito anteriormente para detectar envolturas hidrofóbicas. La sustitución de dicha molécula de agua por el ligando es particularmente favorable si el átomo dador o aceptor del ligando consigue su complemento total de enlaces de hidrógeno formando el enlace de hidrógeno objetivo individual con el grupo de proteínas en cuestión, de modo que la satisfacción de enlaces de hidrógeno adicionales no sea un problema.

Un ejemplo de un grupo adecuado es un nitrógeno plano en un anillo aromático, que se une por ejemplo a un grupo de la cadena principal N-H de la proteína. Un ejemplo ilustrativo de una estructura este tipo, que se sepa que es esencial para conseguir una alta potencia experimentalmente, se muestra en la Figura 5; el sistema representado es SB220025 que se une a p38 MAP quinasa. Se sabe que el enlace de hidrógeno Met 109 es importante para la potencia, ya que son enlaces de hidrógeno análogos en otras quinanas. El término de la función de puntuación que se ha descrito anteriormente permite que dichos enlaces de hidrógeno se detecten automáticamente, y antes de la medida experimental. De igual importancia, los falsos positivos, que comparten superficialmente algunas características del motivo estructural requerido pero que carecen de un componente fundamental, también pueden ser rechazados automáticamente. El rechazo de falsos positivos se ha optimizado mediante la ejecución de una variante dada de la función de puntuación, que identifica ligandos de bases de datos de alta puntuación con enlaces de hidrógeno especiales en ubicaciones no observadas en compuestos activos conocidos, examinando la estructura resultante, y alterando parámetros para eliminar el refuerzo para los casos de ensayo en particular en cuestión.

Un proceso 200 para detectar enlace de hidrógeno sencillo en un entorno hidrofóbico incluye lo siguiente. El átomo dador o aceptor debe estar en un anillo con la excepción del nitrógeno, que se permite que esté en un átomo que no está en el anillo. Si el átomo del ligando es un grupo dador, entonces todos los otros dadores del grupo (por ejemplo, dos hidrógenos de NH₂) pueden estar unidos por H a la proteína. Sólo los átomos de proteínas de la estructura principal pueden participar en este tipo de enlace de H especial. El átomo dador o aceptor de proteína en cuestión en la estructura de apoproteína (es decir, la estructura sin ligando presente) debe tener menos de 3 moléculas de agua desolvatación en la primera envoltura. El número de moléculas de agua en la primera envoltura se determina usando un programa basado en la retícula para la adición de moléculas de agua una estructura de proteína o complejo proteína-ligando (código de adición de agua). Si estos criterios se cumplen, la suma del factor angular de la puntuación de empaquetado hidrofóbico se hace sumando el átomo pesado unido a H del ligando y los átomos de carbono unidos a este átomo del ligando. En término empaquetamiento hidrofóbico especializado usado en esta suma sólo contiene el peso angular de la puntuación del empaquetamiento hidrofóbico que se describe en el proceso 100 y no el peso de contacto. Si la suma del empaquetamiento fóbico absoluto está por encima de un punto de corte, se considera que el enlace de H está en un entorno hidrofóticamente restringido y se aplica un refuerzo de enlace de H especial de -1,5 kcal/m a este enlace de H. En algunos casos, se encuentra que una pequeña perturbación del ligando puede mover el término de empaquetamiento hidrofóbico especializado por encima del punto de corte. Por lo tanto, si el término está inicialmente por debajo del punto de corte, pequeñas perturbaciones del cuerpo rígido del ligando se forman en una magnitud de 0,3 Å. En cada geometría perturbada, el término se vuelve a calcular y si en alguna geometría el término supera el punto de corte, se aplica un refuerzo. Esto ayuda a evitar las discontinuidades inherentes en el uso de un punto de corte.

El proceso 200 identifica una estructura en la que se debería asignar un refuerzo "especial" a un enlace de hidrógeno sencillo. Una segunda situación se produce cuando existen múltiples enlaces de hidrógeno correlacionados entre la proteína y el ligando. La organización de moléculas de agua para solvatar eficazmente una estructura de este tipo en la geometría confinada del sitio activo puede ser incluso más problemática que la situación que se ha descrito anteriormente. Esto se produce solamente si las moléculas de agua implicadas en dicha solvatación están en un entorno hidrofóbico exigente, con grupos hidrofóbicos en los dos lados. El acoplamiento de la identificación del enlace de hidrógeno especial con el motivo de envoltura hidrofóbica es crítico si se van a rechazar falsos positivos. Los enlaces de hidrógeno correlacionados se forman rutinariamente en el acoplamiento por ejemplo con pares de la estructura principal muy expuestos al disolvente, pero no existe evidencia a partir de los datos experimentales de que dichas estructuras contribuyan a una mayor potencia en cualquier serie conocida de química medicinal. Detalles en cuanto a la asignación de refuerzos especiales para sistemas con dos o tres enlaces de hidrógeno correlacionados se describen en el Apéndice II.

Los enlaces de hidrógeno múltiple se definen con respecto al ligando como un par de átomos dador/aceptor, dador/dador, o aceptor/aceptor (mencionado como 'par atómico del ligando'), que se separan por no más de 1 enlace de rotación (los grupos OH cuenta como de no rotación en este cálculo). Varias restricciones sobre los tipos de pares que se puede considerar se hacen tal como se muestra en la siguiente tabla.

1) Evitar enlaces de H de poca calidad (< 0,05 para la puntuación del par HB)
2) No incluir pares que implican el mismo átomo de proteína neutra,
3) Evitar pares implicados en un puente salino si el potencial electrostático en cada átomo del ligando no está por debajo de un punto de corte.
4) Evitar pares del puente salino si el átomo de proteína está implicado en un puente salino proteína-proteína.
5) Evitar pares dador/dador del ligando que vienen de NH _x x > = 2 grupos. N que no está en el anillo y con carga formal cero.
6) Evitar átomos de proteína cargados formalmente con más de 8 moléculas de agua de envoltura secundaria en el estado de no unión.

7) Evitar enlaces de H cargado-neutro a menos que el átomo de la proteína esté en un puente salino.
8) Evitar pares de diferentes átomos aceptores neutros en el ligando.
9) Los pares de enlace de H neutros deben satisfacer comprobaciones del entorno del átomo del anillo y de fobicidad.
10) Evitar que OH de ligandos se una a H de proteínas si el átomo de la proteína tiene carga formal cero.

Si los átomos del ligando del par individualmente tienen carga formal neta cero, entonces deben satisfacer la siguiente comprobación de hidrofobicidad y criterios para conseguir un refuerzo de enlace de H especial. En primer lugar, los átomos del ligando deben ser parte del mismo anillo o pueden ser átomos que no están en el anillo sólo si se conectan directamente con el mismo anillo. Suponiendo que el par satisface estas restricciones, la hidrofobicidad de la región del enlace de H se detecta de una forma similar a la del enlace sencillo de H. Una suma de la puntuación de empaquetamiento hidrofóbico que se ha descrito anteriormente se hace usando el par de átomos pesados del ligando del enlace de H y los átomos en el anillo conectados directamente a los átomos del par del ligando. Si un átomo del ligando del par no es un átomo del anillo pero está conectado a un anillo, la suma incluye átomos del anillo que son vecinos cercanos al átomo del ligando que no está en el anillo. Si la suma fóbica absoluta descrita está por encima de un punto de corte, se da un refuerzo especial de 3 kcal/m al par unido a H. El recuento doble de enlaces de pares y de H especial sencillo se evita buscando pares primero y excluyendo cualquier enlace de H reforzado encontrado en la lista de pares a partir de la lista de refuerzo del enlace de H sencillo.

Los refuerzos del enlace de H especial disminuyen linealmente con la calidad del enlace de H. La calidad del enlace de H se mide en el sentido de la puntuación del enlace del H del par usando la distancia dador/aceptor y el ángulo formado por el vector dador-átomo-pesado -H y el vector aceptor de H. Para átomos del aceptor del ligando en anillos se hace una comprobación para la extensión a la que el vector del par libre de electrones del aceptor está alineado con el vector dador-átomo- pesado -H. Una descripción detallada de cómo se consigue esto matemáticamente se proporciona en el Apéndice 2.

Hemos identificado un número considerable de complejos proteína-ligando en los que están presentes los motivos que contienen dos o tres enlaces de hidrógeno correlacionados, que satisfacen los criterios anteriores (incluyendo el requisito de envoltura hidrofóbica). Un número de ejemplos se muestran a continuación. El primero, que se muestra en la Figura 6, representa estaurosporina unida a la CDK2. Este tipo de par correlacionado se encuentra también en un número de otras quinasas. Algunas de las CDK2 activas, tales como cgp60474, forman tres enlaces de hidrógeno correlacionados; esta estructura se muestra en la Figura 7.

La Figura 8 muestra un segundo ejemplo con estreptavidina unida a biotina. Aquí se producen tres enlaces de hidrógeno correlacionados en una región envuelta hidrofóbicamente. Hasta el momento actual, ninguna función de puntuación empírica ha explicado la afinidad de unión excepcionalmente grande de la estreptavidina en biotina. Una vez que se reconoce el motivo de enlace de hidrógeno envuelto, triple y se asigna una puntuación apropiada, la desviación entre la afinidad de unión calculada y experimental, usando una estructura acoplada, es solamente 1,5 kcal/mol.

Un tercer ejemplo diferente, que se muestra en la Figura 9, es un inhibidor de fiderestat unido a la estructura lef3 de la aldosa reductasa, que también presenta un enlace de hidrógeno correlacionado triple. La gran puntuación adicional asociada con este enlace de hidrógeno es necesaria para explicar por qué este compuesto destaca (baja afinidad de unión nanomolar experimentalmente) de entre el gran número de ligandos que consiguen grandes puntuaciones hidrofóbicas (en un experimento de identificación sistemática virtual) en el bolsillo altamente hidrofóbico del complejo nativo lef3.

La combinación de la envoltura hidrofóbica con 1 a 3 enlaces de hidrógeno posicionados correctamente es característica de cada enlace de hidrógeno neutro-neutro "especial" que hemos identificado en nuestros estudios de diversas dianas farmacéuticamente relevantes. Sin embargo, se requieren características adicionales para eliminar falsos positivos. En particular, si la pareja del enlace de hidrógeno en la proteína está muy expuesta a disolvente, entonces la formación de una estructura capaz de solvatar la estructura en cuestión, mientras que aún se permite que las moléculas de agua implicadas formen un número adecuado de enlaces de hidrógeno adicionales, se hace más fácil. Por lo tanto, necesitamos que el grupo o grupos de la proteína implicados en el enlace o enlaces de hidrógeno especiales tengan un número limitado de moléculas de agua en la primera o segunda envoltura. El número de dichas moléculas de agua se determina usando el método de adición de agua que se ha analizado anteriormente.

La magnitud de los refuerzos asociados con los enlaces de hidrógeno especiales se ha determinado por optimización frente a una gran base de datos experimentales, que contiene un número significativo de ejemplos de cada tipo de estructura. Los valores se proporcionan en la tabla que sigue a continuación.

Resto de enlace de H	Refuerzo
Enlace sencillo con el anillo en entornos fóbicos	1,5 kc/m
Par Neutro en entornos fóbicos	3,0 kc/m

Hemos identificado las siguientes características que mejoran la señal de enlaces de hidrógeno cargado-cargado.

(1) Número de moléculas de agua que rodean al componente de proteína del puente salino. Es poco probable que grupos cargados que están totalmente expuestos al disolvente participen en el enlace de hidrógeno cargado-cargado. El esfuerzo para desplazar el disolvente es demasiado grande.

(2) Número de enlaces de hidrógeno cargado-cargado formados por el grupo del ligando cargado. Hemos observado que existen tres tipos diferentes de estructuras formadas por puente salino:

(a) Monodentada (enlace de hidrógeno sencillo) entre el grupo del ligando y un grupo de proteína.

(b) Bidentado (dos enlaces de hidrógeno) entre el grupo del ligando y un grupo de proteína. Un ejemplo de un puente salino bidentado en la trombina, entre un grupo amino cargado positivamente y un carboxilato incrustado en el bolsillo de especificidad apropiada, se muestra en la Figura 10.

(c) Enlaces de hidrógeno de un grupo del ligando a dos grupos diferentes de proteínas. Esto requiere tener dos grupos de proteínas de tipo cargado en proximidad cercana. Este tipo de estructura, que supuestamente crea energía de deformación en la proteína, se produce frecuentemente. Parece que es un tipo especial de motivo de reconocimiento molecular que se desarrolla mediante la evolución de la presión.

La Figura 11 ilustra un ejemplo que muestra el ligando Gr217029 que se une al término N9 de la versión del virus de la gripe (pdb código 1bji) del receptor de neuramididasa. La distancia entre los oxígenos del carboxilato es solamente 4,5 Å en este caso. Tanto los datos empíricos como el razonamiento físico químico sugieren que enlaces de hidrógeno del tipo (c), mencionado anteriormente, proporcionan más energía de estabilización que los de (b), que a su vez proporciona más energía de estabilización que los de (a). A menos que la condición (1) se satisfaga apropiadamente, no es probable que ninguna de las tres estructuras sea favorable esta de energía libre. Es la combinación del acceso restringido de agua para el grupo de proteínas, una interacción electrostáticas excepcionalmente fuerte entre el ligando y el grupo de proteína, lo que crea un motivo de reconocimiento molecular.

(3) Para zwitteriones, una razón principal por la que un valor por defecto para enlaces de hidrógeno cargado-cargado se establezca en cero es que, en la formación de un puente salino, tanto la proteína como el ligando deben entregar contribuciones de largo alcance a la energía de Born (es decir, más allá de la primera envoltura). Hacer el complemento de la primera envoltura de enlaces de hidrógeno es bastante posible cuando se forma un puente salino, pero el reemplazo del volumen de agua con la proteína, o moléculas de agua unidas, reduce claramente la respuesta dialéctica posible al ión. Para un ión monovalente, el campo de Coulomb disminuye aproximadamente a $1/r$, mientras que los efectos de largo alcance (al menos hasta la segunda envoltura) hacen una contribución no trivial a la energía libre de solvatación total. Sin embargo, para un zwitterión, los campos de las cargas positivas y negativas cancelan hasta cierto punto la distancia larga (la cantidad depende de la separación de los dos grupos cargados), produciendo un campo dipolar para el que las contribuciones de la segunda envoltura (y más elevada) a la energía libre de solvatación se reducen significativamente. Por lo tanto, la formación de dos puentes salinos por el zwitterión, particularmente si los dos restos cargados están próximos espacialmente, debería ser más favorable que en la unión a un solo ión. Un ejemplo de una unión a zwitterión de este modo se muestra en la Figura 11.

(4) Distinguimos entre casos en los que el ligando esta cargado positivamente y la proteína cargada negativamente, y viceversa.

(5) Fuerza del campo electrostático en el ligando. Para recibir una mayor afinidad de unión por un puente salino, el sitio en el que está ubicada la carga del ligando debe ser electrostáticamente favorable. Sumamos el campo electrostático identificado sistemáticamente en el sitio del ligando, e imponemos puntos de corte para asignar refuerzos basándose en la optimización empírica sobre nuestra serie de casos de ensayo. Estos puntos de corte ayudan a reducir el número de falsos positivos que reciben refuerzos de carga-carga especiales.

En la siguiente tabla se enumeran los diversos refuerzos de carga-carga especiales, tanto con respecto a criterios (basándose en las cinco categorías que se han analizado anteriormente) como valores numéricos asignados a los refuerzos. Sus últimos se han optimizado basándose en el ajuste de todo nuestra serie de ensayos.

Interacción de Cargas	Refuerzo
Átomo del ligando cargado en entornos de bajo potencial electrostático.	1,5 kcal/m
Configuración del zwitterión, intervalo de refuerzos que aumentan con la atracción electrostática	3,0 a 4,7 kcal/m
Grupo de ligando positivo que se une al grupo negativo de proteína débilmente solvatado.	0,5 kcal/m
Ligando co2- grupo H unido a múltiples restos positivos de proteína positiva cercanos	1,0 kcal/m
Par del puente salino en entornos de baja solvatación (menos de 9 moléculas de agua de segunda envoltura sobre un átomo de la proteína con carga pre-ligada)	2,0 kcal/m

La siguiente tabla muestra las puntuaciones activas de la función de puntuación para una serie de receptores de Glu2 en comparación con las energías de unión experimentales.

5

Ligando	Nuestra puntuación	ΔG _(kc/m)
1ftl	8,1	8,3
1pwr	10,5	13,0
1ftj	9,3	8,5
1mm7	11,0	12,7
1mqj	12,0	11,9
1ftm	10,5	8,8
1n0t	7,0	5,7
1ftl	8,1	8,3
1m5b	10,8	9,3
1m5c	9,5	9,3
1m5e	9,1	6,3

El receptor de Glu2 es un sistema, junto con la neuramididasa, por el que las interacciones electrostáticas son particularmente importantes, y por lo tanto proporciona una importante contribución al conjunto de entrenamiento para estos tipos de interacciones.

10

Se ha observado empíricamente que los ligandos más grandes tienden a ser favorecidos por la mayoría de los algoritmos de puntuación. Hemos aplicado dos tipos de términos basándose en estas observaciones:

15

(1) Se asigna un refuerzo a ligandos más pequeños basándose en su masa. El término que usamos es lineal y cubre el intervalo de masas de 300 a 450. Por debajo de 300, todos los ligandos reciben el refuerzo máximo (1 kcal/mol); por encima de 450, todos los ligandos reciben 0.

20

(2) Los términos que reconocen apilamiento aromático, e interacciones de catión pi, de los que hemos planteado hipótesis durante muchos años sobre que desempeñan un papel significativo en las afinidades de unión proteína-ligando para complejos específicos que poseen uno o ambos de estos motivos. Hemos incluido varios sistemas en esta categoría en nuestro conjunto de datos (por ejemplo, fxa, acetilcolinesterasa) y desarrollado una descripción geométrica de cada tipo de interacción, optimizada para incluir compuestos activos, y eliminar falsos positivos. Los valores para las diversas interacciones se asignaron en el curso de los estudios de enriquecimiento y afinidad de unión.

25

Otros dos términos de puntuación se derivan de un 'mapa del sitio' de una superficie fóbica/polar de la proteína. En primer término 'fóbico polar' de atracción refuerza átomos polares en regiones fóbicas del mapa del sitio activo. Las regiones fóbicas se determinan en este método como puntos de una red con interacciones de vdW favorables y campos eléctricos bajos de la proteína. Los átomos polares son átomos no aceptores no dadores que no son carbono. Este término de atracción por lo general está en orden de 0,5 kcal/m por lo que no contribuye significativamente al promedio. Una puntuación de 'repulsión fóbica' del mapa del sitio de repulsión penaliza átomos del ligando fóbicos encontrados en regiones de la proteína con campos eléctricos elevados. Se supone que estas regiones en este método están ocupadas con agua. Se terminó tan en contribuye al nivel de 0,3 kcal/m como promedio y con menos frecuencia que el término fóbico polar.

35

La siguiente ecuación incorpora todos los términos usados para calcular los componentes de nuestra función de puntuación que favorecen la unión. Hemos optimizado estos términos basándonos en afinidades de unión experimentales y estructuras cristalinas.

$$E_{unión} = E_{fo_emp} + E_{electro} + E_{HB_espec} + E_{HB_pr} + E_{fo_pr} + E_{PI}$$

5

en la que E_{fo_emp} es el empaquetamiento fóbico/término de envoltura; $E_{electro}$ contiene cualquiera de los refuerzos electrostáticos; E_{HB_espec} contiene refuerzos para enlaces de H en regiones fóbicas; E_{HB_pr} es el término del par de enlace de H convencional; E_{PI} contiene el catión pi y términos de apilamiento de pi; y finalmente E_{fo_pr} es el término del par lipofílico convencional. Tal como se ha indicado anteriormente, este término se debe complementar con términos de desolvatación y entrópicos para calcular la puntuación final para la afinidad de unión, que se indica a continuación. La eficacia de nuestro método se basa en el cálculo de afinidades de unión para un conjunto de complejos proteína-ligando en el PDB, usando fórmulas acopladas (mostrando de este modo que se pueden obtener buenos resultados para la predicción de la afinidad de unión con fórmulas acopladas en oposición a las estructuras experimentales). La tabla que sigue a continuación presenta resultados para nuestra función de puntuación para las RMSD estructurales a partir de afinidades de unión de los dichos complejos analizados en el presente documento, en la que "Puntuación XP" se refiere a resultados obtenidos a partir de nuestra función de puntuación y la "Puntuación SP" representa puntuaciones que usan una función de puntuación más antigua para la comparación. Las afinidades de unión experimentales se incluyen en la tabla cuando estén disponibles.

20

Entrada en PDB	Puntuación XP	ΔG_{exp}	XP RMS	Puntuación SP	ΔG_{exp}	SP RMS
1aaq	-11,1	-10,8	12,1*	-12	-10,8	1,6
1abe	-8,9	-9,2	0,3	-11,5	-9,2	0,4
1abf	-6,1	-7,2	0,1	-11,8	-7,2	0,1
1acj	-9,8	-10	2,8	-7,6	-10	4,6
1acm	-13,7	--	0,4	-15,4	--	0,3
1aco	-4,2	--	0,3	-9,7	--	1
1add	-11,1	-9,2	0,5	-9,3	-9,2	0,6
1adf	-9,1	-6,2	4,5	-13,7	-6,2	2,9
1aha	-8,6	--	0,2	-7,9	--	0,1
1ake	-22	--	1,3	-10,3	--	2,8
1apb	-7,8	-7,9	0,1	-11,6	-7,9	0,1
1apt	-12,3	-12,8	0,9	-11	-12,8	0,6
1apu	-9,8	-10,5	1,3	-8,8	-10,5	1
1apv	-11,7	-11,3	0,7	-9,8	-11,3	0,4
1apw	-11,2	-10,9	1	-9,6	-10,9	0,3
1atl	-10,6	--	0,9	-8,2	--	2,4
1avd	-17,2	--	0,6	-10,2	--	0,5
1azm	-5,1	--	1,9	-6,3	--	1,5
1b6j	-12,1	--	2,8	-16	-	3,1
1b6k	-14,6	--	1,3	-13,9	--	0,7
1b6l	-10,5	--	1,1	-11,2	--	0,6
1b6m	-14	--	0,2	-12,7	--	3,2
1baf	-4,2	--	1,2	-9,3	--	1
1bap	-8,9	-9,4	0,3	-11,6	-9,4	0,4
1bbp	-15,2	--	5,1	-13	--	5,1

ES 2 432 753 T3

1bkm	-11,1	--	3,6	-15,4	--	2
1bma	-6,2	--	0,6	-7,9	--	2,1
1bra	-5,4	-2,5	2,3	-7,2	-2,5	0,3
1byb	-14,6	--	0,5	-16,4	--	0,5
1c1b	-15,5	--	0,9	-9,6	--	0,5
1c3i	-12,7	--	0,6	-11,7	--	0,4
1c5p	-5,5	--	0,5	-7,3	--	0,2
1c83	-8,5	--	0,3	-10	--	0,1
1c84	-10,9	--	0,3	-9,6	--	0,3
1c86	-11,2	--	0,2	-11,3	--	0,2
1c87	-10,5	--	0,4	-11,7	--	0,3
1c88	-10,7	--	0,2	-12,2	--	0,2
1c8k	-7,3	--	3,6	-7,5	--	5,5
1cbs	-7,5	--	0,5	-6,6	--	0,4
1cbx	-9,8	-8,6	0,3	-13,4	-8,6	0,5
1cde	-15,1	--	2,3	-10,6	--	2,1
1cdg	-5,6	--	6,9	-5,7	--	6,5
1cil	-5,6	--	3,7	-6,8	--	3,9
1cnx	-6,2	--	6,7	-9,2	--	6,6
1com	-9,2	--	3,7	-8,6	--	3,7
1coy	-8,8	--	0,4	-8,7	--	0,3
1ctr	-7,8	-6,8	1	-6,4	-6,8	2,4
1ctt	-6,3	-6,2	0,6	-6,5	-6,2	5
1d3d	-15,7	--	2,1	-10,7	--	2,6
1d3p	-13,2	--	1,8	-10,8	--	2,1
1d7x	-7,2	--	0,5	-9,7	--	0,4
1d8f	-7,7	--	9,3	-9	--	4,8
1dbb	-13,2	-12,3	0,3	-9,4	-12,3	0,4
1dbj	-13,2	-10,4	0,3	-8,6	-10,4	0,3
1dbk	-12,7	-11	0,9	-8,1	-11	0,4
1dbm	-12,6	-12,9	0,5	-9,1	-12,9	2
1dd6	-14	--	8,1	-11	--	8,1
1dds	-9,8	-11,3	3,6	-8,1	-11,3	2,4
1dhf	-9,5	-10,1	5,8	-9,3	-10,1	5,6
1did	-4,7	--	3,7	-6,9	--	4,1
1die	-5,9	-2,9	0,4	-7,7	-2,9	0,8
1dih	-9,5	-7,8	2,4	-15,8	-7,8	4
1dm2	-12,2	--	0,7	-9,5	--	0,7
1dog	-7	-5,5	3,7	-10,3	-5,5	3,7
1drl	-4,7	-5,57	0,3	-7,8	-5,57	1,5
1dwb	-5,6	-4	0,6	-6,7	-4	0,3

ES 2 432 753 T3

1dwc	-10	-10,3	2,1	-10,5	-10,3	0,8
1dwd	-13,7	-11,4	1,4	-11,5	-11,4	1,4
1e5i	-9,9	--	0,3	-10,8	--	0,2
1eap	-12,1	--	0,8	-10,2	--	2,4
1ebg	-13	-14,8	0,4	-20	-14,8	0,3
1ecv	-8,2	--	0,2	-9,8	--	0,2
1eed	-10,2	-6,7	11,6	-10,7	-6,7	10,8
1ejn	-9,3	--	0,4	-9,4	--	0,3
1ela	-8,1	-8,7	2	-5,7	-8,7	10
1elb	-2,2	--	4,1	-7,6	--	5,2
1elc	-7,3	-9,8	7,5	-6,3	-9,8	7,8
1eld	-8,5	--	0,8	-4,7	--	6,7
1ele	-8,1	--	0,4	-6,8	--	0,3
1epb	-9,7	--	2,6	-6,5	--	2,3
1eta	-3,8	--	1,2	-4,1	--	1,7
1etr	-11,5	-10,1	0,7	-9,5	-10,1	0,7
1ets	-13	-11,6	1,3	-11,8	-11,6	1,4
1ett	-12,7	-8,4	0,6	-9,5	-8,4	0,5
1ezq	-16,2	--	1	-12,9	--	0,2
1f0r	-11	--	2,4	-9,3	--	0,8
1f0s	-11,1	--	1,7	-8,5	--	0,3
1f0t	-10,4	--	1,6	-10	--	0,3
1f0u	-8,6	--	0,6	-11,2	--	1,6
1fen	-10,8	--	1,2	-6,5	--	0,4
1fh8	-9,4	--	0,2	-12,5	--	0,2
1fh9	-7,8	--	0,8	-8,4	--	6,5
1fhd	-9	--	5,4	-11,3	--	0,4
1fjs	-7,9	--	2	-12,6	--	2,4
1fkg	-10,4	--	1,4	-8,2	--	1,6
1fki	-10,7	--	1,3	-7,6	--	1,3
1fq5	-17,8	--	1,9	-15,9	--	2,4
1fvt	-12,1	--	0,9	-8,7	--	0,9
1nco	-8	--	0,6	-11,6	--	0,4
1nis	-4,9	--	0,3	-11,4	--	0,5
1nnb	-5,4	-7,2	1,4	-9,3	-7,2	0,3
1nsc	-5,8	-4,1	0,7	-12	-4,1	0,2
1nsd	-7,2	-7,2	1,1	-10	-7,2	0,3
1odw	-11	--	3,9	-8,7	--	1,5
1okl	-7,8	--	0,5	-5,9	--	2,7
1pbd	-11,3	--	0,3	-9,4	--	0,3
1pgp	-8,3	-7,8	2,3	-10,3	-7,8	2,2

ES 2 432 753 T3

1pha	-12,5	--	0,7	-8,6	--	0,5
1phd	-7,6	--	0,8	-6,2	--	0,3
1phf	-7,9	-6	1,8	-5,3	-6	1,1
1phg	-11,2	-11,8	1,7	-7,7	-11,8	1,2
1poc	-10,7	--	2,5	-12,4	--	1,5
1ppc	-9,1	-8,8	1,5	-10,3	-8,8	1,1
1pph	-8,6	-8,5	0,9	-9,9	-8,5	0,6
1ppi	-15,9	--	1	-13,9	--	7,6
1ppk	-8,5	-10,4	0,6	-11,2	-10,4	0,3
1ppl	-9,7	-11,5	0,9	-12,2	-11,5	2,6
1ppm	-9,3	-7,9	0,5	-12,9	-7,9	0,6
1pro	-15,6	-15,4	1,4	-12,5	-15,4	1,9
1pso	-12,2	-14,1	2,6	-9,8	-14,1	5,7
1sbg	-11,1	-10,6	0,9	-10,8	-10,6	0,4
1slt	-5,2	--	0,7	-9,1	--	0,6
1snc	-9,8	-6,7	2,5	-11,2	-6,7	1,2
1sre	-4,3	-5,4	0,3	-9,6	-5,4	0,3
1srj	-4,8	--	0,4	-9,1	--	0,4
1stp	-18,5	-18,3	0,5	-11,2	-18,3	0,6
1tdb	-8,1	--	7,4	-6,9	--	7,5
1thy	-5,2	--	2,3	-7,4	--	2,5
1tka	-11,2	--	1,7	-15,2	--	2,3
1tlp	-11,5	-10,3	10,4	-13,3	-10,3	7,3
1tmn	-10,2	-10	2,9	-12,7	-10	2,8
1tng	-6,5	-4	0,2	-7,3	-4	0,2
1tnh	-5,4	-4,6	0,6	-7,6	-4,6	0,3
1tni	-3,6	-2,3	2	-5,9	-2,3	2
1tnj	-4,1	-2,7	0,5	-6,5	-2,7	0,4
1tnk	-3,7	-2	1,1	-6,2	-2	1
1tnl	-4,1	-2,6	0,4	-6,2	-2,6	0,2
1tph	-5,2	--	0,2	-8,2	--	0,2
1tpp	-7	--	0,3	-7,7	--	1,1
1trk	-14,6	--	0,9	-14,4	--	2,1
1tyl	-5,1	--	5,2	-5,3	--	1,1
1ukz	-13,5	--	0,5	-12,6	--	0,4
1ulb	-6,4	-7,2	0,3	-9	-7,2	0,4
1wap	-8,1	--	0,2	-9,9	--	0,2
1xid	-6,3	--	4	-6,4	--	4,3
1xie	-5,3	--	2,6	-6,7	--	3,9
2ack	-7,6	--	1,1	-5,8	--	0,9
2ada	-8,8	--	0,4	-10,4	--	0,5

ES 2 432 753 T3

2ak3	-9,7	-5,3	0,8	-8,5	-5,3	0,6
2cgr	-11	-9,9	0,6	-11,3	-9,9	0,5
2cht	-9,9	--	0,4	-10,8	--	0,5
2cmd	-8,6	--	0,8	-10	--	0,5
2cpp	-8,5	-8,3	0,2	-5,8	-8,3	3
2ctc	-7,3	-5,3	1,5	-10,7	-5,3	1,6
2dbl	-12,2	-11,8	2	-8,9	-11,8	0,8
2 gbp	-10,5	-10,4	0,2	-12,9	-10,4	0,1
2ifb	-6,1	-7,4	2,3	-3,5	-7,4	1,9
2lgs	-5,4	--	1	-7,1	--	5,9
2mcp	-4	-7,1	1,5	-5,4	-7,1	1,3
2phh	-10,1	-6,4	0,4	-8,6	-6,4	0,4
2pk4	-6,4	-4,32	0,8	-6,2	-4,32	0,9
2plv	-11,9	--	1,6	-8,2	--	1,2
2r04	-7,4	-6,22	1,1	-7,4	-6,22	0,7
2r07	-8,7	--	1,5	-7,4	--	0,7
2sim	-7,6	--	0,9	-10,6	--	0,8
2tmn	-11	-6,7	0,6	-12,1	-6,7	0,6
2tpi	-8,7	-5,9	0,4	-8,1	-5,9	1,2
2upj	-11,2	-14,2	3,1	-11,3	-14,2	2,7
2xis	-6	-7,9	2,3	-8,1	-7,9	2,4
2yhx	-4,3	--	1,9	-9,6	--	2,2
3cla	-6,5	-4,94	8,5	-4,8	-4,94	6,1
3dfr	-13,8	-14	0,7	-13	-14	0,8
3hvt	-8,8	--	0,7	-8,1	--	0,8
3mth	-5,4	--	1,2	-4,6	--	5,6
3ptb	-6,1	-6,5	0,3	-7,4	-6,5	0,2
3tpi	-7,3	-5,9	0,6	-9,4	-5,9	0,5
4aah	-11,3	--	0,3	-13,3	--	0,3
4cts	-7,7	-	0,2	-8,6	--	0,2
4dfr	-12,7	-13,2	1,1	-12,1	-13,2	1
4fab	-16,4	-14,4	4,5	-8,8	-14,4	0,8
4fbp	-11,7	--	2	-13,9	--	0,5
4fxn	-11,2	--	0,7	-16,6	--	1,2
4hmg	-6,7	-3,5	0,7	-7	-3,5	0,7
4phv	-14,2	-12,5	0,8	-10,2	-12,5	5,5
4tim	-0,8	-2,16	0,8	-11,3	-2,16	1,3
4tln	-5,7	-5,1	1,7	-9,8	-5,1	2,5
4tmn	-12,2	-13,9	1,1	-14,4	-13,9	1,2
4tpi	-6,7	-4	0,7	-8,3	-4	0,6
4ts1	-8,6	-7,6	0,9	-7,5	-7,6	0,8

ES 2 432 753 T3

5abp	-8,1	-9	0,1	-12,1	-9	0,2
5cpp	-8,2	-8	0,1	-5,9	-8	2,7
5cts	-0,2	--	0,2	-8,1	--	0,3
5p2p	-9,6	--	1,9	-9,2	--	10,4
5tim	-1,5	-3,1	0,9	-7	-3,1	0,7
5tln	-11,6	-8,7	2,4	-12,1	-8,7	1,8
5tmn	-10,9	-11	2,6	-13,3	-11	2,8
6abp	-8,9	-9,1	0,3	-11,3	-9,1	0,4
6cpa	-11,4	-15,7	4,8	-14,9	-15,7	4,2
6rnt	-5,5	--	0,6	-8,2	--	0,6
6tim	-5,5	-3,21	0,6	-8,8	-3,21	0,4
6tmn	-11,1	-6,9	2,4	-13,6	-6,9	2,3
7abp	-6,2	-8,1	0,2	-11,7	-8,1	0,2
7cpa	-15,9	-19,1	3,9	-15,2	-19,1	4,1
7cpp	-6,3	-5,2	2	-4,6	-5,2	3,2
7tim	-5,4	--	0,2	-11	--	0,2
8abp	-7,9	-10	0,2	-12,4	-10	0,2
8atc	-11,1	--	0,3	-11,6	--	1,4
8 qch	-7,1	--	0,5	-9,6	--	0,3
9abp	-8	-10,9	0,1	-12,7	-10,9	0,2
9hvp	-10,8	-11,4	1,4	-13,7	-11,4	1,6
1 g45	-8,2	--	7,9	-6,8	--	4,2
1 g46	-8,2	--	8,1	-6,9	--	4,5
1 g48	-6,9	--	2	-7,1	--	3,8
1 g4j	-6,9	--	5,3	-8	--	3,5
1 g4o	-7,9	--	7,9	-6,6	--	4,3
1 g52	-8,6	--	8	-6,7	--	4,3
1 g53	-8,2	--	7,9	-6,4	--	4,5
1 g54	-6,5	--	6	-7,1	--	5,1
1 ghb	-6,8	--	0,3	-8,3	--	0,3
1 glp	-4,2	--	0,8	-10,8	--	0,3
1 q1q	-8,2	--	1,2	-11,7	--	0,3
1 qsp	-7,5	--	1	-9,1	--	2,8
1hbv	-10,2	-8,7	10,5*	-9,9	-8,7	3,2
1hdc	-9,9	--	0,7	-7,8	--	0,5
1hdy	-3,7	--	3,5	-4	--	1,7
1hef	-10,1	-12,1	6,1	-11,4	-12,1	6,8
1hfc	-9,2	--	2,4	-10,3	--	2,3
1hgg	-6,8	--	1,5	-8,7	--	1,4
1hgh	-6,2	--	0,6	-5,7	--	2,3
1hgi	-6,4	--	0,6	-7,7	--	0,2
1hqj	-4,8	--	0,3	-6,4	--	0,4
1hih	-12,5	-11,3	1,4	-11,4	-11,3	1,2

ES 2 432 753 T3

1hps	-12	-12,6	12,1*	-12,8	-12,6	12,1
1hpv	-9,8	-12,6	0,9	-10,3	-12,6	0,9
1hpx	-12,1	-11,2	2,1	-12,1	-11,2	3,5
1hri	-11,3	--	10,5*	-6	--	3,9
1hsq	-14,3	-12,8	3,8	-13,7	-12,8	0,4
1hsl	-9,8	-7,3	1,3	-9,8	-7,3	1,3
1hte	-10,1	-8,6	7,3*	-8,7	-8,6	1,2
1htf	-11,2	-11,1	3,3	-10,1	-11,1	2,2
1hti	-4,3	--	4,4	-6	--	1,7
1hvr	-13,4	-13	1,6	-5,2	-13	3,9
1hyt	-8,8	--	2,6	-10,5	--	0,9
1icn	-7,4	--	9,3	-1	--	8,9
1ida	-9,9	--	2	-12,6	--	12,1
1igj	-6,8	--	0,7	-7,5	--	0,5
1imb	-6,8	--	1,8	-11	--	1,6
1ivb	-6,2	--	0,5	-6,8	--	0,5
1ivc	-5,1	--	1,8	-6,4	--	1,9
1ivd	-3,5	--	0,8	-6,8	--	0,7
1ive	-6,1	--	4,8	-6,4	--	5
1ivf	-8,2	--	0,7	-8,3	--	1,3
1lah	-5,9	--	0,2	-10,7	--	0,1
1lcp	-8,3	--	1,8	-8,9	--	1
1ldm	-7,4	-7,4	1,3	-6,4	-7,4	1,3
1lic	-4,8	--	4,8	-2,8	--	2,4
1lmo	-6,8	--	7,1	-8	--	0,9
1lna	-6,4	--	1,3	-8,2	--	0,7
1lst	-6,3	--	0,7	-8,6	--	0,3
1mbi	-3,9	-2,6	1,9	-3,5	-2,6	1,6
1mcr	-10,1	--	6	-7,6	--	1,7
1mdr	-8,5	--	1,7	-9,2	--	0,5
1mfe	-8,8	-7,2	6,2*	-8,4	-7,2	1,8
1mld	-8,5	--	0,3	-10,2	--	0,2
1mmq	-9,7	--	0,6	-11,4	--	0,3
1mnc	-9,8	-12,3	0,3	-11,4	-12,3	0,4
1mrg	-6,2	--	0,2	-7,5	--	0,1
1mrk	-13,3	--	12	-10	--	1,2
1mup	-5,2	--	4,9	-4,9	--	4,1

5 La RMSD media de la puntuación XP calculada para las afinidades de unión experimentales que usan nuestra función de puntuación es 1,8 kcal/mol, que representa una mejora de aproximadamente dos veces sobre la RMSD de puntuación SP de 3,3 kcal/mol para el mismo conjunto de datos. Indicamos que el cálculo de la RMSD que se ha analizado anteriormente incorpora estimaciones de la energía de tensión del receptor, lo que conduce a una pequeña modificación de los errores que se podrían calcular a partir de la tabla anterior para algunos. Hay muy pocos casos que usan nuestra función de puntuación con errores superiores a 3 kcal/mol, lo que implica que la presente función de puntuación tiene una capacidad excelente para distinguir enlazadores débiles (milimolar),

moderados (micromolar), y fuertes (nanomolar) entre sí, una tarea principal en la identificación sistemática virtual, una capacidad que sólo está presente ligeramente en las funciones de puntuación anteriores. Para la mejora más allá de este nivel con respecto a la precisión, creemos que se debe introducir flexibilidad en el receptor, y se debe incorporar un nivel adicional de detalle con respecto a las interacciones proteína-ligando.

La eficacia de nuestro método se basa también en el acoplamiento de un conjunto de compuestos activos conocidos para un receptor en particular en ese receptor, y además en el acoplamiento de una base de datos de 1.000 moléculas de tipo fármaco elegidas aleatoriamente como un conjunto de comparación. Calculamos el "enriquecimiento" de compuestos activos de nuestra nueva función de puntuación (que se define a continuación) en comparación con ligandos de bases de datos, y comparamos los resultados con varias funciones de puntuación menos eficaces, más antiguas.

Nuestro conjunto de ensayos solamente incluye ligandos que se ajustan en la versión particular del receptor en estudio. Las formas del receptor se eligen de modo que están relativamente "abiertas" para maximizar el número de ligandos que se ajustan. Indicamos la fracción de ligandos que hemos conservado de un conjunto de datos iniciales que hemos investigado. Los ligandos se han descartado principalmente basándose en oposiciones estéricas que bloquean el acceso a un bolsillo. Cualquier ligando que consigue un modo de unión razonable (evaluado mediante comparación directa con una estructura experimental, o por analogía con una estructura conocida) se ha conservado independientemente de la puntuación que se le asigna.

En el cálculo de la puntuación de enriquecimiento solo se usan ligandos que tienen afinidades de unión experimentales superiores a 10 μM . Las afinidades y de unión de algunos ligandos que se unen con menos fuerza que éste se calculan como una comprobación de la capacidad de nuestra función de puntuación para discriminar ligandos que se unen débilmente y fuertemente. El punto de corte en 10 μM se elige por que es lo que se usa por lo general para definir compuestos "activos" en experimentos de identificación sistemática de alto rendimiento. El porcentaje de ligandos aleatorios de bases de datos de los que se espera que compitan en este intervalo de actividad (suponiendo una evaluación perfecta computacional y experimental de la afinidad de unión) por lo general es del orden de un 0,1-1,0 % (aunque estos números pueden ser significativamente superiores o inferiores dependiendo del receptor) o 1-10 compuestos en una base de datos de 1.000 compuestos elegidos aleatoriamente. Por lo tanto, sería esperar que positivos ciertos en la base de datos aleatorios introdujeran relativamente poco ruido en nuestros datos de enriquecimiento. La siguiente tabla ilustra compuestos activos totales con 10 μM o mejor afinidad de unión en comparación con el total que se ajusta y acopla en el receptor elegido para cada identificación sistemática.

Identificación sistemática	Compuestos activos totales	Total de pocillos acoplados
Trombina	15	15
HIV-RT	32	24
COX-2	19	17
P38 1kv2	10	10
P38 b17	50	37
EGFR-tk	127	110
CDK-2	10	6
TK	7	6
Fxa	20	16
Estrógeno	10	8
LCK	169	121
Neuramida.	9	9
AchE	28	25
HIV-proteasa	15	14

Las definiciones convencionales de enriquecimiento encuentran problemas de encuentro cuando se usa un gran número de ligandos en un conjunto de ensayos con respecto a las bases de datos aleatorios. Usamos una definición que es independiente del número de compuestos activos en el conjunto del ensayo, ya que está determinado solamente por el número de ligandos aleatorios de la base de datos que puntúan mejor que cada activo (por lo tanto, la clasificación de compuestos activos de unos con respecto a otros elimina como un elemento).

La siguiente tabla presenta un número promedio de ligandos de la base de datos que supera al conjunto que compuestos activos conocidos para los receptores que hemos estudiado, para tres funciones de puntuación diferentes; la nueva función de puntuación, que contiene los términos habilitados mediante nuestros métodos que se han descrito anteriormente; y dos de puntuación más antiguas, que se sabe que son competitivas con otros enfoques en la bibliografía. Los compuestos activos tienen al menos una actividad de 10 μ M. El promedio necesario se calcula determinando el número de ligandos de la base de datos que destaca sobre cada activo, añadiendo a continuación estos valores en conjunto y dividiendo por el número de compuestos activos. Por lo tanto, un valor de cero introducido en la tabla que sigue a continuación representa una puntuación perfecta, en la que todos los compuestos activos superan a los demás ligandos de la base de datos. Se puede observar que, con esta medida, la función de puntuación actual es cualitativamente mejor que las funciones de puntuación más antiguas para muchos (sino la mayoría de los receptores) a menudo por un factor de 10 o más.

Identificación Sistemática	Señuelos que Superan el Promedio			
	Función de puntuación	Puntuación Actual Función de puntuación	Antigua 1 Función de puntuación	Antigua 2 Función de puntuación
AchE		80	560	389
Neuramididasa		14	523	113
Factor XA		5	179	46
P38 1kv2		1	25	187
P38 1bl7		16	133	278
HIV-RT		15	46	110
COX-2		18	57	121
CDK-2		2	18	180
Trombina		1	76	17
HIV proteasa		22	124	60
Estrógeno		11	4	25
LCK quinasa		35	317	291
EGFR		32	407	452
Termolisina		5	N/D	0
Timidina Quinasa		0	0	13

Los métodos que se han descrito anteriormente se pueden usar de muchas formas. En un ejemplo en particular, tal como se ha descrito, los métodos se pueden usar para calcular valores numéricos para contribuciones de diversos términos a la afinidad de unión. En otro ejemplo en particular, los contribuyentes fundamentales para cada término se pueden ilustrar visualmente en una interfaz gráfica de usuario (GUI). La visualización de dichos términos es útil en el diseño de nuevas moléculas, que podría aprovechar las características fundamentales de la cavidad del sitio activo de la proteína (tal como sitios en los que las moléculas de agua están envueltas hidrofólicamente tal como se ha descrito anteriormente). También se puede usar una visualización para determinar las regiones de un compuesto activo que no son importantes para conseguir una unión fuerte a la proteína. Dichas regiones las puede modificar a continuación los químicos médicos para dar a la molécula otras propiedades que son necesarias para una eficacia farmacéutica *in vivo*, por ejemplo, mejor absorción a través del sistema digestivo, evitación de toxicidad, mayor solubilidad, y mejora de reacciones metabólicas no favorables.

La envoltura hidrofóbica se puede visualizar identificando el grupo de átomos hidrofóbicos en el ligando que están envueltos, identificando los átomos o restos de la proteína cuya participación en la envoltura se detecta con el algoritmo, y usando forma, color, y textura de la presentación de los átomos del ligando y de la proteína en cuestión para resaltar que estas regiones.

Esta visualización puede permitir que un químico médico entienda qué partes del ligando deben permanecer hidrofóbicas (es decir, no se deberían modificar para incorporar un grupo hidrofílico), y que restos/átomos de la proteína son responsables de la creación de la envoltura. También es posible usar datos a partir de múltiples ligandos para crear un mapa completo de regiones de la proteína que son capaces de conducir a un término de envoltura favorable. A continuación se puede diseñar un nuevo ligando que aproveche tantas de estas regiones como sea posible, y que aproveche cada región en la mayor medida.

- 5 Los enlaces de hidrógeno neutro-neutro que obtienen refuerzos se pueden visualizar resaltando la proteína y los grupos dadores y aceptores del ligando que forman los enlaces de hidrógeno en cuestión, así como los grupos de proteína que contribuyen a la envoltura hidrofóbica requerida del enlace de hidrógeno. Dicha visualización indicaría grupos fundamentales en el ligando que solo se deberían reemplazar con otros grupos capaces de conseguir los mismos refuerzos especiales.
- 10 Los enlaces de hidrógeno cargado-cargado especiales se pueden visualizar resaltando la proteína y los grupos dadores y aceptores del ligando que forman el enlace o enlaces de hidrógeno en cuestión.
- 15 Para ambos enlaces de hidrógeno neutro-neutro y cargado-cargado, existe un criterio para asignar el refuerzo que implica evaluar el grado de solvatación del grupo en cuestión en la proteína. Esto se puede visualizar mediante la visualización de moléculas de agua explícitas que rodean los grupos en cuestión.
- 20 La visualización puede ser útil incluso si las contribuciones numéricas a la afinidad de unión no se calculan. La visualización puede hacer una contribución específica al diseño de nuevos compuestos, y dar lugar a procesos de optimización en los que los compuestos activos se modifican para mejorar una amplia gama de propiedades al mismo tiempo que se conserva la potencia.
- 25 Las realizaciones de la invención se pueden aplicar en circuitos electrónicos digitales (o en hardware, firmware, software del ordenador), o en combinaciones de los mismos. Las realizaciones de la invención se pueden aplicar como un producto para programa informático, es decir, un programa informático incorporado de forma tangible en un soporte de información (por ejemplo, en un dispositivo de almacenamiento legible con máquina o en una señal propagada) para su ejecución por, o para controlar el funcionamiento de, aparatos de procesamiento de datos (por ejemplo, un procesador programable, un ordenador, o múltiples ordenadores). Un programa informático se puede escribir con cualquier forma de lenguaje de programación, que incluye los lenguajes compilado o interpretado, y que se puede desplegar de cualquier forma, que incluye como un programa autónomo o como un módulo, componente, subrutina, u otra unidad adecuada para su uso en un entorno informático. Un programa informático se puede desplegar para su ejecución en un ordenador o un múltiples ordenadores en un sitio o distribuidos a través de múltiples sitios e interconectados mediante una red de comunicación.
- 30 Las etapas del método de las realizaciones de la invención se pueden realizar con uno o más procesadores programables que ejecutan un programa informático para realizar funciones de la invención funcionando sobre datos de entrada y generando datos de salida. Las etapas del método también se pueden realizar con, y el aparato de la invención se puede aplicar como, circuitos lógicos con fines especiales (por ejemplo, una FPGA (matriz de puertas programables por campo) o un ASIC (circuito integrado de aplicación específica).
- 35 Los procesadores adecuados para la ejecución de un programa informático incluyen, a modo de ejemplo, microprocesadores de fines tanto generales como especiales, y uno cualquiera o más procesadores de cualquier tipo de ordenador digital. Generalmente, un procesador recibirá instrucciones y datos de una memoria de sólo lectura o una memoria de acceso aleatorio o ambas. Los elementos esenciales de un ordenador son un procesador para ejecutar instrucciones y uno o más dispositivos de memoria para almacenar instrucciones y datos. Generalmente, un ordenador también incluirá, o estará acoplado operativamente para recibir datos de o transferir datos a, o ambos, uno o más dispositivos de almacenamiento masivo para almacenamiento de datos (por ejemplo, discos magnéticos, discos magneto ópticos, o discos ópticos). Los soportes de información adecuados para contener instrucciones y datos del programa informático incluyen todas las formas de memoria no volátil, que incluyen a modo de ejemplo dispositivos de memoria de semiconductores (por ejemplo, EPROM, EEPROM, dispositivos de memoria ultrarrápida; discos magnéticos (por ejemplo, discos duros internos o discos extraíbles; discos magneto ópticos; y discos CD ROM DVD-ROM). El procesador en la memoria se puede complementar con, o incorporar en circuitos lógicos con fines especiales.
- 40 La invención se puede aplicar en un sistema informático que incluye un componente de respaldo (por ejemplo, como un servidor de datos) o que incluye un componente middleware (por ejemplo, un servidor de aplicaciones) o que incluye un componente frontal (por ejemplo, un ordenador cliente que tiene una interfaz gráfica de usuario o un navegador Web a través del cual un usuario puede interactuar con una aplicación de la invención) o cualquier combinación de dichos componentes de respaldo, middleware, o frontal. Los componentes del sistema pueden estar interconectados mediante cualquier forma o medio interconectar de comunicación de datos digitales (por ejemplo, una red de comunicación). Ejemplos redes de comunicación incluyen una red de área local ("LAN") y una red de área amplia ("WAN") (por ejemplo, Internet).
- 45 El sistema informático puede incluir clientes y servidores. Un cliente y un servidor generalmente están alejados entre sí y por lo general interactúan a través de una red de comunicación. La relación entre cliente y servidor se presenta en virtud de programas informáticos que se ejecutan en los respectivos ordenadores y que tienen una relación cliente-servidor entre sí.
- 50 Se debe observar que la descripción anterior pretende ilustrar to y no limitar el alcance de la invención, que se define por el alcance de las reivindicaciones adjuntas. Otras realizaciones están dentro del alcance de las siguientes
- 55
- 60
- 65

reivindicaciones.

Apéndice I

5 Puntuación de Empaquetamiento Hidrofóbico

1) Asignar los grupos hidrofóbicos de ligandos:

10 Los átomos de ligandos se asignan a 'grupos fóbicos' N_g cada uno con un conjunto de átomos $\{j\}$. Cada grupo fóbico se define por un conjunto de átomos de carbono conectados, sin interrumpir mediante conexiones atómicas polares. Átomos de carbono conectados a átomos polares no están incluidos en el grupo. Grupos de tamaño uno no contribuyen a la puntuación.

15 A partir de este punto para cada grupo se hace la puntuación siendo el resultado neto la suma de las puntuaciones de cada grupo. Lo siguiente (etapas 2 a través de 6) describe la puntuación para cualquier grupo en particular.

2) Calcular los contactos con átomos de proteínas fóbicas;

20 Para cada átomo del ligando j que está en el grupo fóbico, calcular el número total de contactos, $\text{cont}(j)$ con átomos de proteínas fóbicas que están dentro de una distancia de $3,0 \text{ \AA}$ más la suma de los radios de vdw del átomo del ligando (vdw_j) y de la proteína (vdw_k) (la suma total es aproximadamente 6 \AA para átomos típicos de ligando y proteína). Para cada par de contacto c_{jk} que corresponde a pares de átomos (j, k), almacenar la distancia del contacto $r_{jk}(j, c_{jk})$ y el vector $v(j, c_{jk}) = r_k - r_j$ entre los átomos de ligando y proteína en contacto. Tal como se indicaba anteriormente $r_{jk} \leq 3,0 + \text{vdw}_i + \text{vdw}_j$. En este proceso, mantener el seguimiento del par de contacto jk_{min} con la distancia de contacto más pequeña r_{jk} . Marcar el vector del par de contacto más pequeño v_0 . Este vector se denomina 'vector de referencia'.

3) Calcular los ángulos con el vector de referencia.

30 Para cada átomo del ligando j en el grupo y todos sus contactos $c_{jk} = (1, \text{cont}(j))$, calcular el ángulo, $\text{ang}(j, c_{jk})$ entre el vector $v(j, c_{jk})$ y el de vector referencia v_0 .

4) Calcular la puntuación de contacto radial.

35 Para un átomo dado j de un grupo, la puntuación del contacto radial $S_j(r)$ es una función de rampa lineal entre 0 y 1 derivada como sigue a continuación a partir de los contactos totales que hace el ligando, $\text{cont}(j)$.

$$\begin{aligned} S_j(r) &= 0,0 \text{ si } \text{cont}(j) \leq 20 \\ S_j(r) &= 1,0 \text{ si } \text{cont}(j) \geq 32 \\ S_j(r) &= (\text{cont}(j) - 20,0)/12,0 \text{ si } 20 < \text{cont}(j) < 32. \end{aligned} \quad (\text{A1})$$

40 5) Calcular la función angular.

45 Para un átomo dado j del grupo de cada contacto c_{jk} ($c_{jk} = 1, \text{cont}(j)$) que hace el átomo, se calcula una puntuación angular, $S_{jk}(a)$. Para calcular esta puntuación primero es necesario validar la distancia de contacto $r_{jk}(j, c_{jk})$ en valores enteros $m = 1-6$ por simple tratamiento de la distancia de valor real. La función angular es entonces una función de rampa lineal del ángulo $\text{ang}(j, c_{jk})$ con límites determinados por la validación del contacto m ;

$$\begin{aligned} S_{jk}(a) &= 0,0 \text{ si } \text{ang}(j, c_{jk}) < \text{amín}(m) \\ S_{jk}(a) &= 1,0 \text{ si } \text{ang}(j, c_{jk}) > \text{amáx}(m) \\ S_{jk}(a) &= p(m) * (\text{ang}(j, c_{jk}) - \text{amín}(m)) / (\text{amáx}(m) - \text{amín}(m)). \end{aligned} \quad (\text{A2})$$

50 siendo $\text{amáx}(m)$ y $\text{amín}(m)$ los ángulos máximo y mínimo por validación m tal como sigue a continuación (en grados)

$m = 1-3:$ $\text{amín} = 120, \text{amáx} = 140$

$m = 4$ $\text{amín} = 120, \text{amáx} = 140$

$m = 5$ $\text{amín} = 140, \text{amáx} = 180$

$m = 6$ $\text{amín} = 150, \text{amáx} = 180$

y p(m) es el factor de peso para cada intervalo angular:

$$\begin{aligned} m = 1-4; & \quad p = 1,0 \\ m = 5 & \quad p = 0,7 \\ m = 6 & \quad p = 0,6 \end{aligned}$$

5 La puntuación angular total $S_j(a)$ para el átomo j se calcula a continuación simplemente como la suma de todos los valores $S_{jk}(a)$;

$$S_j(a) = \sum_{k=1, \text{cont}(j)} S_{jk}(a) \quad (A4)$$

10 6) Calcular puntuación total del grupo, puntuación de empaquetamiento total.

La puntuación del empaquetamiento total del grupo para el grupo ith de $S_g(i)$ es simplemente la suma de las puntuaciones angulares $S_j(a)$ multiplicada por las puntuaciones del contacto $S_i(r)$;

$$S_g(i) = \sum_j S_j(r) * S_j(a). \quad (A5)$$

15 La puntuación del empaquetamiento total hidrofóbico E_{fo_emp} es la suma negativa de las puntuaciones del grupo;

$$E_{fo_emp} = -\sum_{i=1, N_g} S_g(i). \quad (A6)$$

20 Apéndice II

Escalado de refuerzos de enlaces de H especiales

25 Lo siguiente describe el escalado de la puntuación del enlace de H especial ($E_{HB \text{ espec}}$) a medida que la geometría difiere de la idealidad.

30 Un enlace de H típico se produce entre un 'dador' de hidrógeno (enlace de NH por ejemplo), y un 'aceptor de hidrógeno' (átomo de O del carbonilo). La calidad del enlace de H se mide por el grado en el que el ángulo θ entre el vector del enlace N-H y el vector H - O difiere de la colinealidad (un ángulo de 180 grados es colineal). Además, la distancia R_{OH} entre los átomos de H y O está idealmente en el intervalo de aproximadamente 1,8 Å. La función de puntuación típica del enlace de H usa funciones lineales de los parámetros θ y R_{OH} para designar una calidad de E_{hb} del enlace de H como;

$$E_{hb} = A * S(r) * S(\theta) \quad (A9)$$

35 en la que A es una constante y S(r) mide la desviación de la distancia 'óptima' r_0 ;

$$\begin{aligned} S(r) &= 1,0, \text{ si } R_{OH} \leq r_0 \\ S(r) &= (R_{OH} - r_0)/(r_m - r_0), \text{ si } r_0 < R_{OH} < r_m \\ S(r) &= 0,0, \text{ si } R_{OH} \geq r_m \end{aligned} \quad (A10)$$

r_m es la distancia máxima permisible del enlace de H. La función angular tiene una forma similar:

$$\begin{aligned} S(\theta) &= 1,0, \text{ si } \theta \geq \theta_0 \\ S(\theta) &= (\theta - \theta_m)/(\theta_0 - \theta_m), \text{ si } \theta_m < \theta < \theta_0 \\ S(\theta) &= 0,0, \text{ si } \theta < \theta_m \end{aligned} \quad (A11)$$

40 en la que θ_0 es el ángulo máximo y θ_m es el ángulo mínimo permitido en la parametrización.

45 Hasta que un enlace de H cumpla los requisitos como un par de enlace de H especial, debe tener un valor de E_{hb} superior a -0,05 tal como se parametriza en Glide. Una vez que se declara adicionalmente que el enlace de H especial cumple todos los criterios descritos en la Sec. C, se asigna un refuerzo de base E_{Base} de -3 o -2 (-3 para un par y -2 para el segundo par si existen dos pares que comparten el mismo átomo). Este refuerzo de base se reduce

a escala a continuación usando una función angular de la forma de la Ec. (A11) con θ_0 como 135° y θ_m de 120° . A cada enlace del par se le asigna un factor de escala angular S1, S2. El producto del factor de escala angular para cada enlace $S1*S2$ del par es el factor de escala neta aplicado al refuerzo de base,

$$E_{hb_espec} = S1*S2*E_{base} \quad (A12)$$

5

Si el átomo del ligando está cargado formalmente, no se aplica ningún factor de escala.

- 10 Una segunda clase de enlaces de H bonds implica átomos aceptores en anillos (por lo general átomos de N) con un átomo de H. En este caso la calidad del enlace de H es óptima cuando el eje del dador es paralelo a la dirección del par libre de electrones del dador único. La dirección del par libre de electrones se aproxima como la suma de los vectores que conectan los átomos R1 a A, y R2 a A, en la que R1 y R2 son los átomos del anillo conectados al átomo A aceptor del anillo. Se calcula el ángulo θ entre el vector del par libre de electrones y el eje del dador N-H.
- 15 Un factor de escala S_{1p} de la forma de la Ec. (A11) se calcula con $\theta_0 = 130^\circ$ y $\theta_m = 100^\circ$. En estos casos los factores de escala S1 o S2 se multiplican por S_{1p} , es decir, dos porciones angulares escalan una escala del enlace de H dado del par.

REIVINDICACIONES

1. Un método implementado por ordenador para caracterizar la interacción de un ligando propuesto (20) con un receptor especificado (22) mediante el acceso a datos a partir de los que el ordenador calcula un valor representativo de dicha interacción, VRI, comprendiendo dicho método puntuación por ordenador de interacciones hidrofóbicas entre uno o más átomos del ligando (24-32, 50) y uno o más átomos del receptor (34-38, 56), comprendiendo dicha puntuación por ordenador al menos siguientes etapas:
- a) recibir datos representativos de átomos en el ligando acoplado con el receptor;
 - b) actuar sobre dichos datos para identificar grupos de al menos tres átomos del ligando lipofílico conectados covalentemente que presentan interacción hidrofóbica con uno o más átomos del receptor lipofílico e interacciones hidrofóbicas de puntuación de dichos átomos del ligando con dichos átomos del receptor;
 - c) asignar una primera asignación por la presencia de envoltura hidrofóbica para cada uno de dichos grupos de al menos tres átomos del ligando lipofílico conectados covalentemente que presentan interacción hidrofóbica con átomos del receptor en lados opuestos de dicho grupo, y no para cualquier grupo cuyo tamaño sea inferior a tres; y
 - d) calcular el VRI que incluye dicha primera asignación determinada en la etapa (c).
2. El método de la reivindicación 1 en el que el cálculo de VRI incluye adicionalmente al menos un término representativo de al menos uno de los siguientes: energía libre, entalpía y entropía.
3. El método de la reivindicación 1 en el que las interacciones hidrofóbicas de puntuación incluyen el cálculo de un componente de puntuación del área superficial, SASC, que varía como una función del área superficial del contacto hidrofóbico-hidrofóbico entre átomos del ligando y átomos del receptor, y la primera asignación por la presencia de envoltura hidrofóbica da como resultado un aumento del VRI con respecto a un VRI hipotético que se obtendría usando el SASC pero no la primera asignación.
4. El método de la reivindicación 1 en el que las interacciones hidrofóbicas de puntuación incluyen el cálculo de un término de energía del par átomo-átomo, y la primera asignación por la presencia de envoltura hidrofóbica da como resultado un aumento del VRI con respecto a un VRI hipotético que se obtendría usando la suma del término de energía del par átomo-átomo pero no la primera asignación.
5. El método de la reivindicación 3 en el que las interacciones hidrofóbicas de puntuación incluyen adicionalmente el cálculo de un término de energía del par átomo-átomo, y la primera asignación por la presencia de envoltura hidrofóbica da como resultado un aumento del VRI con respecto a un VRI hipotético se obtendría usando la suma del término de energía del par átomo-átomo más el término SASC, pero no la primera asignación.
6. El método de cualquiera de las reivindicaciones 1-5 en el que dicho valor de VRI se usa con un programa informático y el método se usa para simular el acoplamiento del ligando en el receptor para producir la estructura del complejo receptor-ligando.
7. El método de la reivindicación 6 en el que el método se usa para ayudar en la asignación de una estructura al complejo receptor-ligando.
8. El método de la reivindicación 6 en el que el VRI se usa para evaluar la afinidad de unión del ligando a una estructura del complejo proteína-ligando producida por el programa de acoplamiento.
9. El método de cualquiera de las reivindicaciones 1, 2 o 4 en el que el cálculo de dicho valor de VRI comprende asignar un valor representativo de la interacción receptor-ligando en una estructura predeterminada de receptor-ligando.
10. El método de la reivindicación 9 en el que dicha estructura predeterminada de receptor-ligando se obtiene a partir de un programa de acoplamiento a partir de simulaciones dinámicas moleculares, o a partir de minimizaciones que usan un modelo continuo de solvatación.
11. El método de cualquiera de las reivindicaciones 1, 2 o 4 que comprende adicionalmente generar una presentación gráfica de características estructurales del complejo receptor-ligando que contribuyen a un mayor VRI debido a envoltura hidrofóbica.
12. El método de la reivindicación 11 que comprende adicionalmente producir una presentación gráfica generada por ordenador del complejo en el que los átomos del receptor que contribuyen a la envoltura hidrofóbica de uno o más átomos del ligando están resaltados gráficamente para proporcionar una imagen física de la envoltura.
13. El método de la reivindicación 11 en el que superficies del receptor que están implicadas en la envoltura hidrofóbica del ligando están resaltadas en la presentación.

14. El método de cualquiera de las reivindicaciones 1-4 en el que dicho método se usa para seleccionar compuestos a ensayar experimentalmente.
- 5 15. El método de la reivindicación 1 en donde el método comprende una o más de las siguientes etapas en cualquier orden siempre y cuando la etapa (g) se realice en último lugar:
- 10 a) calcular un término que represente dicha primera asignación de envoltura hidrofóbica;
 b) calcular un término que represente refuerzos electrostáticos;
 c) calcular un término que aumente VRI para enlaces de hidrógeno (H) en regiones hidrofóbicas;
 d) calcular un término que represente pares de enlaces de H;
 e) calcular términos que representen cationes π y apilamientos π ;
 f) calcular un término que represente un par lipofílico convencional; y
 g) calcular la puntuación a partir de una suma de términos.
- 15 16. El método de la reivindicación 1 en el que la presencia de envoltura hidrofóbica se determina realizando al menos las siguientes etapas:
- 20 a) localizar átomos del receptor lipofílico que están dentro de una distancia especificada del punto de corte de un átomo hidrofóbico del ligando;
 b) determinar la distribución espacial y angular de átomos del receptor lipofílico alrededor del átomo del ligando hidrofóbico;
 c) adjudicar la primera asignación basándose en la distribución espacial y angular de los átomos del receptor lipofílico alrededor del átomo del ligando hidrofóbico.
- 25 17. El método de la reivindicación 1 en el que un valor para la primera asignación atribuible a dicho grupo conectado de átomos del ligando se determina realizando al menos las siguientes etapas:
- 30 a) calcular la distancia entre un átomo del ligando lipofílico en el grupo y uno o más átomos lipofílicos del receptor;
 b) evaluar el grado de envoltura del átomo del ligando lipofílico mediante una fórmula que incluye términos que representan la dispersión angular de los átomos del receptor lipofílico y su distancia desde el átomo del ligando lipofílico, y asignar el valor para la primera asignación que se está basando en dicha evaluación del grado de envoltura.
- 35 18. El método de la reivindicación 17 en el que una mayor dispersión angular de átomos del receptor que están dentro de una distancia especificada del punto de corte del átomo del ligando da como resultado un valor de VRI más elevado usando dicha fórmula.
- 40 19. El método de la reivindicación 1 en el que dicho grupo de átomos del ligando conectados incluye un conjunto de átomos de carbono conectados sin interrumpir mediante conexiones de átomos polares.
20. El método de la reivindicación 1 en el que el ligando incluye grupos de átomos del ligando neutros y el método incluye:
- 45 a) identificar uno o más enlaces de hidrógeno entre los átomos neutros del ligando y el receptor;
 b) determinar átomos del ligando conectados al grupo o grupos de enlace de hidrógeno;
 c) asignar una segunda asignación a enlaces de hidrógeno basándose en (a) si existe envoltura hidrofóbica de los átomos del ligando implicados en los enlaces de hidrógeno, y los átomos del ligando conectados a los grupos de enlace de hidrógeno, y además en (b) si los átomos del ligando identificados en (a) se benefician de envoltura en la que (i) los átomos del ligando son lipofílicos o (ii) los átomos del ligando han hecho su complemento total de enlaces de hidrógeno sin la envoltura,
- 50 estando adicionalmente dicho método **caracterizado por que** se detecta un enlace sencillo de hidrógeno del ligando con el receptor, y se adjudica una segunda asignación de enlace de hidrógeno sencillo especial si ese enlace de hidrógeno y su entorno satisfacen una o más de las siguientes condiciones:
- 55 (a) un átomo dador o aceptor de enlaces de hidrógeno del ligando es: i) nitrógeno; ii) en un anillo o iii) ambos (i) y (ii);
 (b) un átomo del ligando forma parte de un grupo de átomos dadores de enlaces de hidrógeno y todos los dadores del grupo están unidos por hidrógeno al receptor;
 (c) el receptor es una proteína y átomos del receptor que participan en el enlace de hidrógeno son parte de la estructura principal de la proteína;
 (d) el átomo dador o aceptor del receptor apropiado en la estructura de apoproteína está poco solvatado;
 (e) se calcula la suma de los factores angulares en términos de envoltura para el átomo pesado del ligando implicado en el enlace de hidrógeno y cualquiera de los átomos de carbono unidos a este átomo, y se requiere que la suma supere un umbral especificado y se adjudica una segunda asignación de enlace de hidrógeno
- 60
65

sencillo especial si ese enlace de hidrógeno y su entorno satisfacen la siguiente condición: la suma de los factores angulares en términos de envoltura para el átomo pesado del ligando implicado en el enlace de hidrógeno y se calcula cualquiera de los átomos de carbono unidos a este átomo, y se requiere que la suma supere un umbral especificado.

5 21. El método de la reivindicación 1 en el que ligando incluye grupos de átomos del ligando neutros y el método incluye:

- 10 a) identificar uno o más enlaces de hidrógeno entre los átomos neutros del ligando y el receptor;
 b) determinar átomos del ligando conectados al grupo o grupos de enlace de hidrógeno;
 c) asignar una segunda asignación a enlaces de hidrógeno basándose en (a) si existe envoltura hidrofóbica de los átomos del ligando implicados en los enlaces de hidrógeno, y los átomos del ligando conectados a los grupos de enlace de hidrógeno, y además en (b) si los átomos del ligando identificados en (a) se benefician de envoltura en la que (i) los átomos del ligando son lipofílicos o (ii) los átomos del ligando han hecho su complemento total de enlaces de hidrógeno sin la envoltura,

15 adicionalmente **caracterizado por que**

la etapa (d) comprende el uso de un programa de adición de agua basado en la retícula para evaluar la solvatación del átomo dador o aceptor en una estructura de apoproteína del receptor.

20 22. El método de la reivindicación 21 en el que se requieren átomos que tengan menos de 3 moléculas de agua en contacto para cumplir con los requisitos para la segunda asignación.

25 23. El método de la reivindicación 1 en el que el ligando incluye grupos de átomos del ligando neutros y la puntuación de interacciones hidrofóbicas incluye:

- 30 a) identificar uno o más enlaces de hidrógeno entre los átomos neutros del ligando y el receptor;
 b) determinar átomos del ligando conectados al grupo o grupos de enlace de hidrógeno;
 c) asignar una segunda asignación a enlaces de hidrógeno basándose en (a) si existe envoltura hidrofóbica de los átomos del ligando implicados en los enlaces de hidrógeno, y los átomos del ligando conectados a los grupos de enlace de hidrógeno, y además en (b) si los átomos del ligando identificados en (a) se benefician de envoltura en la que (i) los átomos del ligando son lipofílicos o (ii) los átomos del ligando han hecho su complemento total de enlaces de hidrógeno sin la envoltura, comprendiendo el método designar pares de enlaces de hidrógeno ligando-receptor y adjudicando una segunda asignación de enlace de hidrógeno del par especial usando uno o más de los siguientes criterios:

- 35 (i) pares de enlaces de hidrógeno múltiples no cumplen con los requisitos para la segunda asignación de enlace de hidrógeno del par especial cuando los átomos de unión están separados por más de un enlace de rotación, en el que los grupos OH cuentan como de no rotación en este cálculo;
 40 (ii) uno o más de los siguientes tipos de pares de enlaces de hidrógeno del receptor de ligando no cumplen con los requisitos para la segunda asignación de enlace de hidrógeno del par especial:

- 45 1) enlaces de H de poca calidad ($< 0,05$ para la puntuación del par HB);
 2) pares que implican el mismo átomo de proteína neutra;
 3) pares implicados en un puente salino si el potencial electrostático en el átomo de ligando no está por debajo de un punto de corte;
 4) pares del puente salino si el átomo del receptor está implicado en un puente salino proteína-proteína;
 5) pares dador/dador del ligando que vienen de grupos NH_x $x \geq 2$ en los que N no forma parte de un anillo y tiene carga formal cero;
 50 6) átomos de proteína cargados formalmente con más de 8 segundas envolturas de moléculas de agua en el estado de no unión;
 7) enlaces de H con carga neutra en donde el átomo de proteína no está en un puente salino;
 8) pares de diferentes átomos aceptores neutros en el ligando;
 9) Pares de enlaces de H neutros que no cumplen individualmente con los requisitos para refuerzos de enlace de hidrógeno sencillo; y
 55 10) el OH del ligando se une al H receptor si el átomo del receptor tiene carga formal cero

60 (iii) para pares en los que los átomos de ligando del par tienen individualmente carga formal neta cero, a los enlaces de hidrógeno se les adjudica una segunda asignación para dicha envoltura hidrofóbica en la que la puntuación de dicha interacción hidrofóbica está optimizada para pares de enlaces de hidrógeno.

24. El método de la reivindicación 23, en el que se usan las siguientes condiciones en la realización de la etapa (c) de la reivindicación 23 para determinar si el par del enlace de hidrógeno tiene una puntuación de envoltura hidrofóbica por encima de una puntuación del punto de corte para recibir la segunda asignación del par del enlace de hidrógeno:

65

- a) átomos del ligando en el par deben formar parte del mismo anillo o, para átomos del ligando que no están en el anillo, los átomos del ligando deben estar conectados directamente al mismo anillo; y
 b) se detecta hidrofobicidad en una región del enlace de hidrógeno y se añaden factores angulares para términos de envoltura y la suma supera dicho umbral especificado usando el par de átomos del ligando del enlace de hidrógeno y los átomos del ligando en el anillo conectados directamente con los átomos del ligando que participan en el enlace de hidrógeno; y
 c) si un átomo del ligando del par no es un átomo del anillo pero está conectado a un anillo, la suma incluye átomos del anillo que son vecinos cercanos al átomo del ligando que no está en el anillo.
25. El método de la reivindicación 23 en el que se evita el doble recuento de enlaces de hidrógeno especiales sencillos buscando primero pares y excluyendo cualquier enlace de hidrógeno reforzado encontrado en la lista de pares a partir de la lista de refuerzo de enlace de H sencillo.
26. El método de la reivindicación 23 en el que se aplica cada uno de los siguientes criterios:
- d) átomos del ligando en el par deben formar parte del mismo anillo o, para átomos del ligando que no están en el anillo, los átomos del ligando deben estar conectados directamente al mismo anillo;
 e) se detecta hidrofobicidad en una región del enlace de hidrógeno y se añaden factores angulares para términos de envoltura tal como se describe en la reivindicación 31, parte (e), usando el par de átomos del ligando del enlace de hidrógeno y los átomos del ligando en el anillo conectados directamente con los átomos del ligando que participan en el enlace de hidrógeno;
 f) si un átomo del ligando del par no es un átomo del anillo pero está conectado a un anillo, la suma incluye átomos del anillo que son vecinos cercanos al átomo del ligando que no está en el anillo; y
 g) la segunda asignación adjudicada al par del enlace de hidrógeno es -3 kcal/mol.
27. El método de la reivindicación 1 en el que el ligando incluye grupos de átomos del ligando neutros y la puntuación de interacciones hidrofóbicas incluye:
- a) identificar uno o más enlaces de hidrógeno entre los átomos neutros del ligando y el receptor;
 b) determinar átomos del ligando conectados al grupo o grupos de enlace de hidrógeno;
 c) asignar una segunda asignación a enlaces de hidrógeno basándose en (a) si existe envoltura hidrofóbica de los átomos del ligando implicados en los enlaces de hidrógeno, y los átomos del ligando conectados a los grupos de enlace de hidrógeno, y además en (b) si los átomos del ligando identificados en (a) se benefician de envoltura en la que (i) los átomos del ligando son lipofílicos o (ii) los átomos del ligando han hecho su complemento total de enlaces de hidrógeno sin la envoltura, y en el que la magnitud de la segunda asignación para envoltura hidrofóbica está evaluada con al menos las siguientes etapas:
- determinar la presencia de singletes de enlaces de hidrógeno, pares correlacionados de enlaces de hidrógeno y tripletes de enlaces de hidrógeno, y
 asignar diferentes valores por enlace a singlete, par y tripletes de enlaces de hidrógeno.
28. El método de acuerdo con la reivindicación 1 en el que el ligando incluye grupos de átomos del ligando neutros y la puntuación de interacciones hidrofóbicas incluye:
- a) identificar uno o más enlaces de hidrógeno entre los átomos neutros del ligando y el receptor;
 b) determinar átomos del ligando conectados al grupo o grupos de enlace de hidrógeno;
 c) asignar una segunda asignación a enlaces de hidrógeno basándose en (a) si existe envoltura hidrofóbica de los átomos del ligando implicados en los enlaces de hidrógeno, y los átomos del ligando conectados a los grupos de enlace de hidrógeno, y además en (b) si los átomos del ligando identificados en (a) se benefician de envoltura en la que (i) los átomos del ligando son lipofílicos o (ii) los átomos del ligando han hecho su complemento total de enlaces de hidrógeno sin la envoltura, y en el que el método que asigna un valor para enlaces de hidrógeno entre grupos de átomos neutros de un ligando y un receptor mediante un método que comprende,
 evaluar el grado de solvatación del receptor, y
 adjudicar un refuerzo de enlace de hidrógeno neutro-neutro dependiendo del grado de solvatación del receptor.
29. El método de la reivindicación 1 en el que el ligando incluye grupos de átomos del ligando neutros y la puntuación de interacciones hidrofóbicas incluye:
- a) identificar uno o más enlaces de hidrógeno entre los átomos neutros del ligando y el receptor;
 b) determinar átomos del ligando conectados al grupo o grupos de enlace de hidrógeno;
 c) asignar una segunda asignación a enlaces de hidrógeno basándose en (a) si existe envoltura hidrofóbica de los átomos del ligando implicados en los enlaces de hidrógeno, y los átomos del ligando conectados a los grupos de enlace de hidrógeno, y además en (b) si los átomos del ligando identificados en (a) se benefician de envoltura en la que (i) los átomos del ligando son lipofílicos o (ii) los átomos del ligando han hecho su complemento total de enlaces de hidrógeno sin la envoltura, y en el que dos pares de enlaces de hidrógeno

satisfacen los criterios para adjudicar una segunda asignación, los dos pares comparten un átomo en común.

30. El método de la reivindicación 29 en el que la segunda asignación para el segundo par es de -2 kcal/mol.

5 31. El método de la reivindicación 23 en el que la segunda asignación asignada al par de los enlaces de hidrógeno, denominada en lo sucesivo "refuerzo de base", se reduce de acuerdo con la siguiente fórmula:

- reducir el refuerzo de base usando una función angular:

$$S(\theta) = (\theta - \theta_m) / (\theta_0 - \theta_m), \text{ si } \theta_m < \theta < \theta_0$$

10

en la que θ_0 es el ángulo máximo y θ_m es el ángulo mínimo permitidos en la parametrización, $S(\theta) = 1,0$, si $\theta \geq \theta_0$ y $S(\theta) = 0,0$, si $\theta < \theta_m$

15

- asignar a cada enlace del par un factor de escala angular S_1, S_2 ; y
- aplicar el producto E_{hb_espec} del factor de escala angular para cada enlace $S_1 * S_2$ del par al refuerzo de base E_{base} , de acuerdo con la fórmula $E_{hb_espec} = S_1 * S_2 * E_{base}$.

32. El método de la reivindicación 31 en el que θ_0 es $135^\circ \pm 5^\circ$ y θ_m es $120^\circ \pm 5^\circ$.

20

33. El método de la reivindicación 28 en el que la solvatación de compañeros del enlace de hidrógeno de la proteína en un enlace de hidrógeno neutro-neutro se evalúa aplicando un algoritmo basado en la red para añadir moléculas de agua explícitas, y hacer el recuento del número de dichas moléculas de agua en la primera y la segunda envolturas alrededor del grupo de proteína en cuestión

25

34. El método de la reivindicación 1 en el que el ligando incluye grupos de átomos del ligando neutros y la puntuación de interacciones hidrofóbicas incluye:

30

- identificar uno o más enlaces de hidrógeno entre los átomos neutros del ligando y el receptor;
- determinar átomos del ligando conectados al grupo o grupos de enlace de hidrógeno;
- asignar una segunda asignación a enlaces de hidrógeno basándose en (a) si existe envoltura hidrofóbica de los átomos del ligando implicados en los enlaces de hidrógeno, y los átomos del ligando conectados a los grupos de enlace de hidrógeno, y además en (b) si los átomos del ligando identificados en (a) se benefician de envoltura en la que (i) los átomos del ligando son lipofílicos o (ii) los átomos del ligando han hecho su complemento total de enlaces de hidrógeno sin la envoltura, y en el que el método comprende adicionalmente producir una representación generada por ordenador del complejo en la que los enlaces de hidrógeno que reciben una segunda asignación especial debido a la envoltura hidrofóbica están resaltados en una presentación gráfica.

35

40

35. El método de la reivindicación 1 en el que el ligando incluye grupos de átomos del ligando neutros y el método incluye:

45

- identificar uno o más enlaces de hidrógeno entre los átomos neutros del ligando y el receptor;
- determinar átomos del ligando conectados al grupo o grupos de enlace de hidrógeno;
- asignar una segunda asignación a enlaces de hidrógeno basándose en (a) si existe envoltura hidrofóbica de los átomos del ligando implicados en los enlaces de hidrógeno, y los átomos del ligando conectados a los grupos de enlace de hidrógeno, y además en (b) si los átomos del ligando identificados en (a) se benefician de envoltura en la que (i) los átomos del ligando son lipofílicos o (ii) los átomos del ligando han hecho su complemento total de enlaces de hidrógeno sin la envoltura, y la envoltura de enlaces de hidrógeno que reciben una segunda asignación especial debido a la envoltura hidrofóbica está resaltada en una presentación gráfica generada por ordenador.

50

36. El método de la reivindicación 35 en el que átomos del receptor que participan en la envoltura hidrofóbica de enlaces de hidrógeno que reciben una segunda asignación están resaltados en una presentación gráfica.

55

37. El método de la reivindicación 35 en el que átomos del receptor que participan en la envoltura hidrofóbica de enlaces de hidrógeno que reciben una segunda asignación están resaltados a través de una presentación de superficies hidrofóbicas del receptor.

60

38. El método de la reivindicación 1 en el que el ligando incluye grupos de átomos del ligando neutros y la puntuación de interacciones hidrofóbicas incluye:

- identificar uno o más enlaces de hidrógeno entre los átomos neutros del ligando y el receptor;
- determinar átomos del ligando conectados al grupo o grupos de enlace de hidrógeno;
- asignar una segunda asignación a enlaces de hidrógeno basándose en (a) si existe envoltura hidrofóbica de

- 5 los átomos del ligando implicados en los enlaces de hidrógeno, y los átomos del ligando conectados a los grupos de enlace de hidrógeno, y además en (b) si los átomos del ligando identificados en (a) se benefician de envoltura en la que (i) los átomos del ligando son lipofílicos o (ii) los átomos del ligando han hecho su complemento total de enlaces de hidrógeno sin la envoltura, y en el que el método comprende adicionalmente producir una representación generada por ordenador del complejo en la que los átomos del ligando conectados a enlaces de hidrógeno que reciben una segunda asignación especial debido a la envoltura hidrofóbica están resaltados en una presentación gráfica.
- 10 39. El método de la reivindicación 1 en el que regiones del receptor capaces de proporcionar envoltura hidrofóbica a grupos adecuados del ligando se detectan realizando simulaciones dinámicas moleculares e identificando regiones en las que las moléculas de agua tienen potenciales químicos desfavorables.
- 15 40. El método de la reivindicación 1 o en el que regiones del receptor capaces de proporcionar envoltura hidrofóbica a grupos adecuados del ligando se detectan colocando moléculas de agua rodeando al receptor y detectando entornos envueltos con estas moléculas de agua.

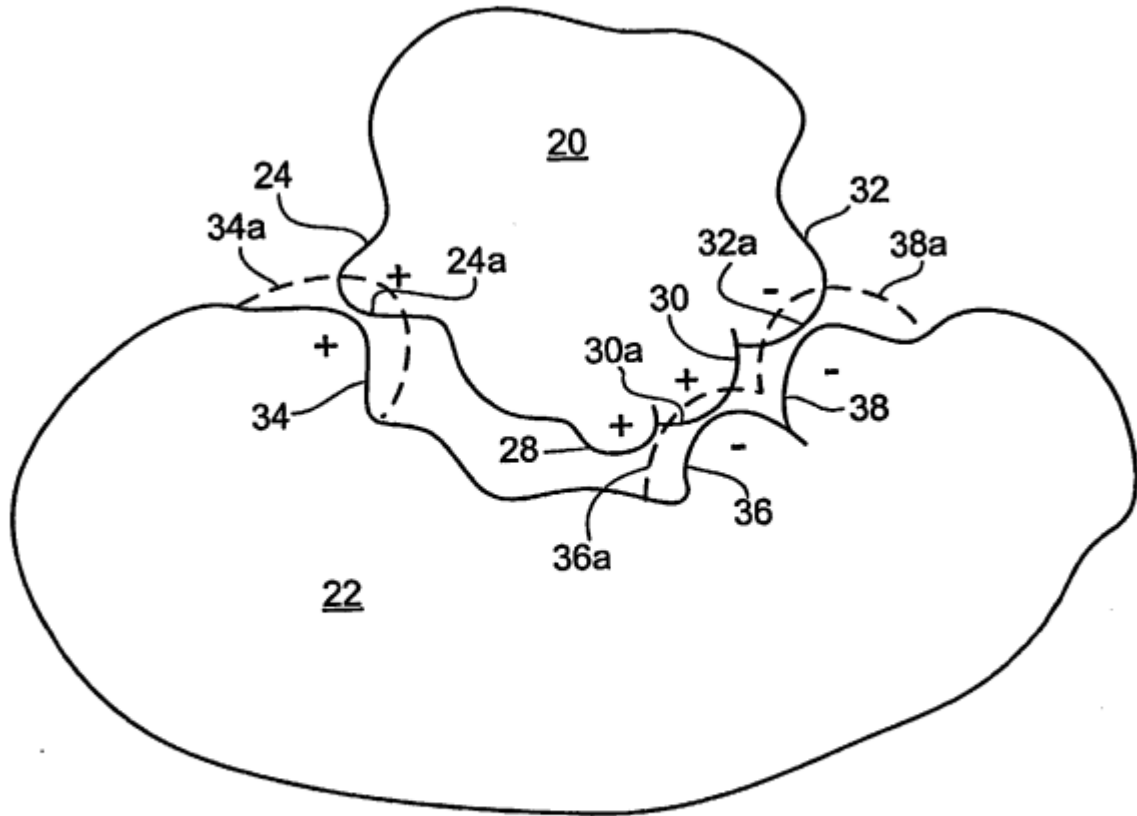


FIG. 1

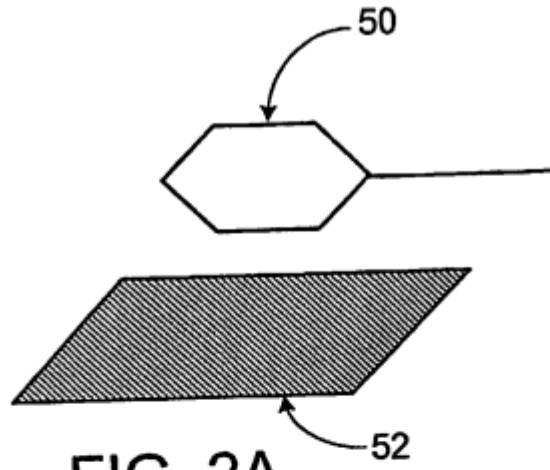


FIG. 2A

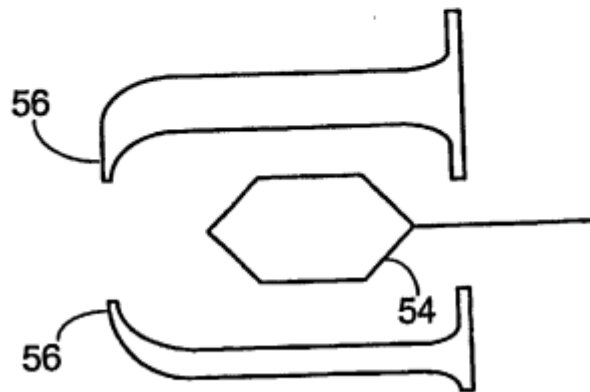


FIG. 2B

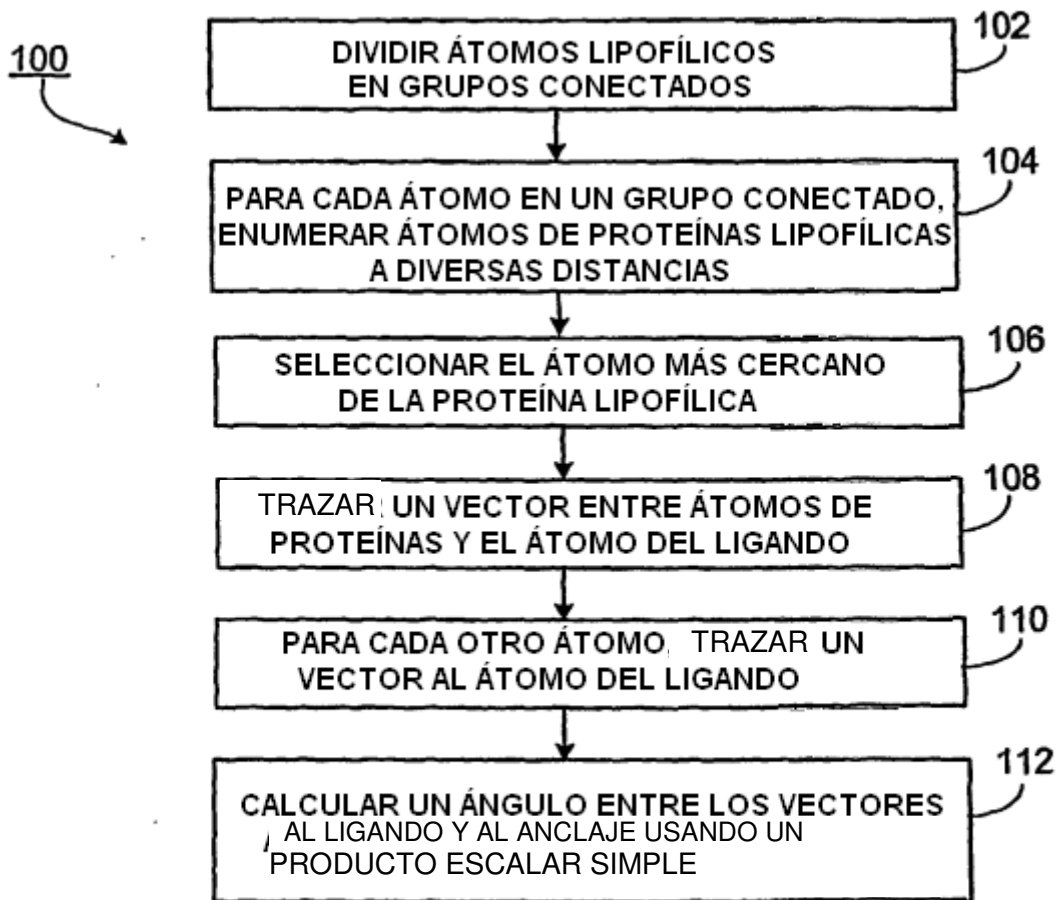
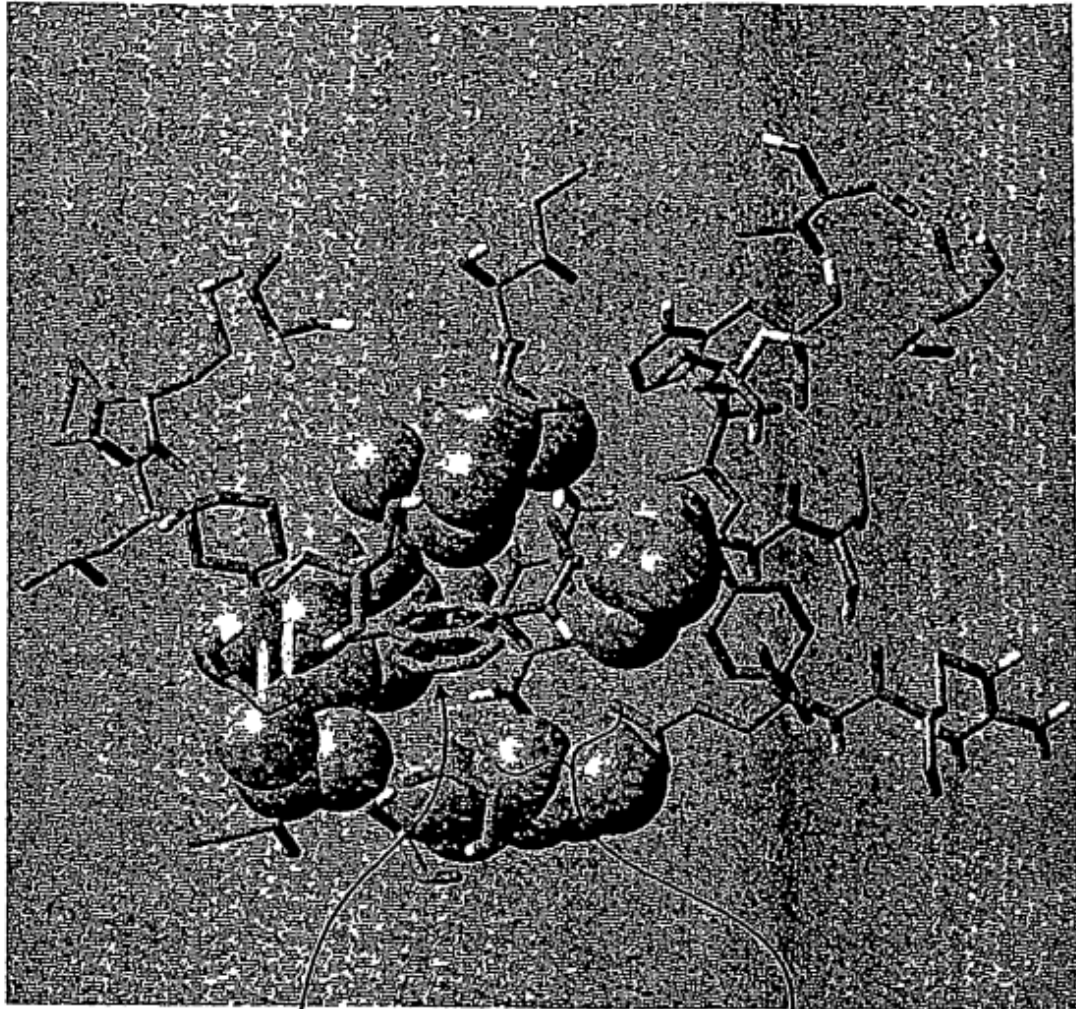


FIG. 3



60

FIG. 4

62

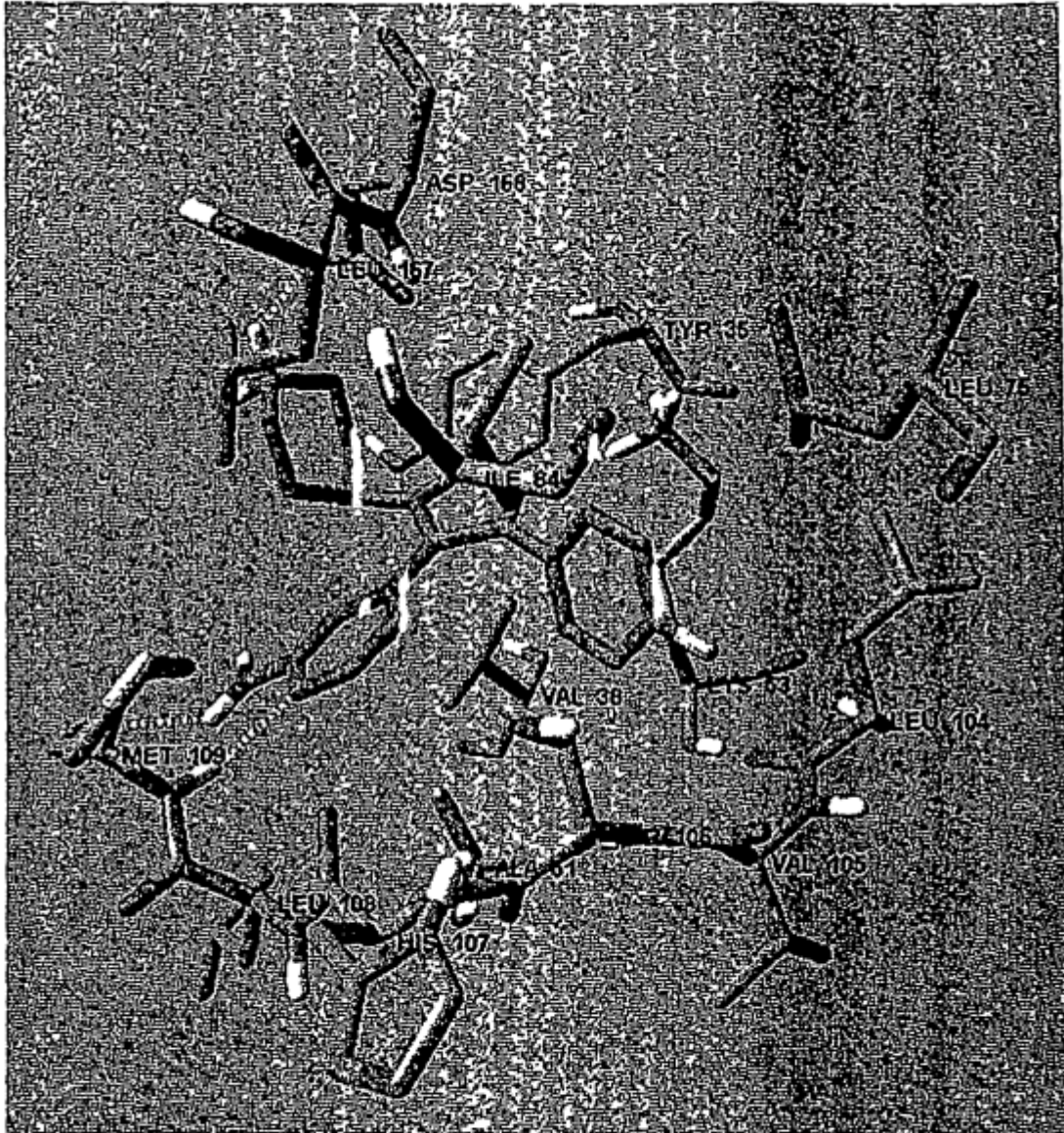


FIG. 5

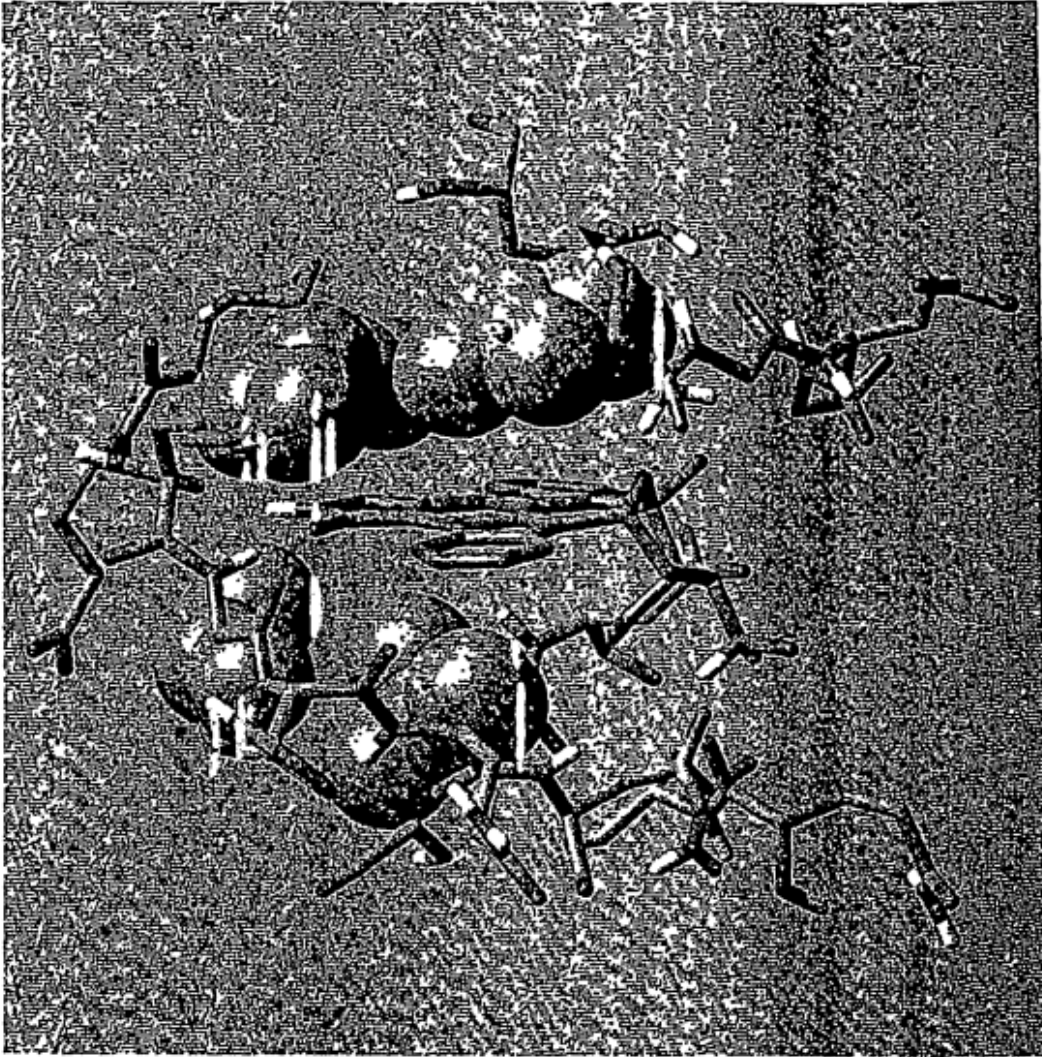


FIG. 6

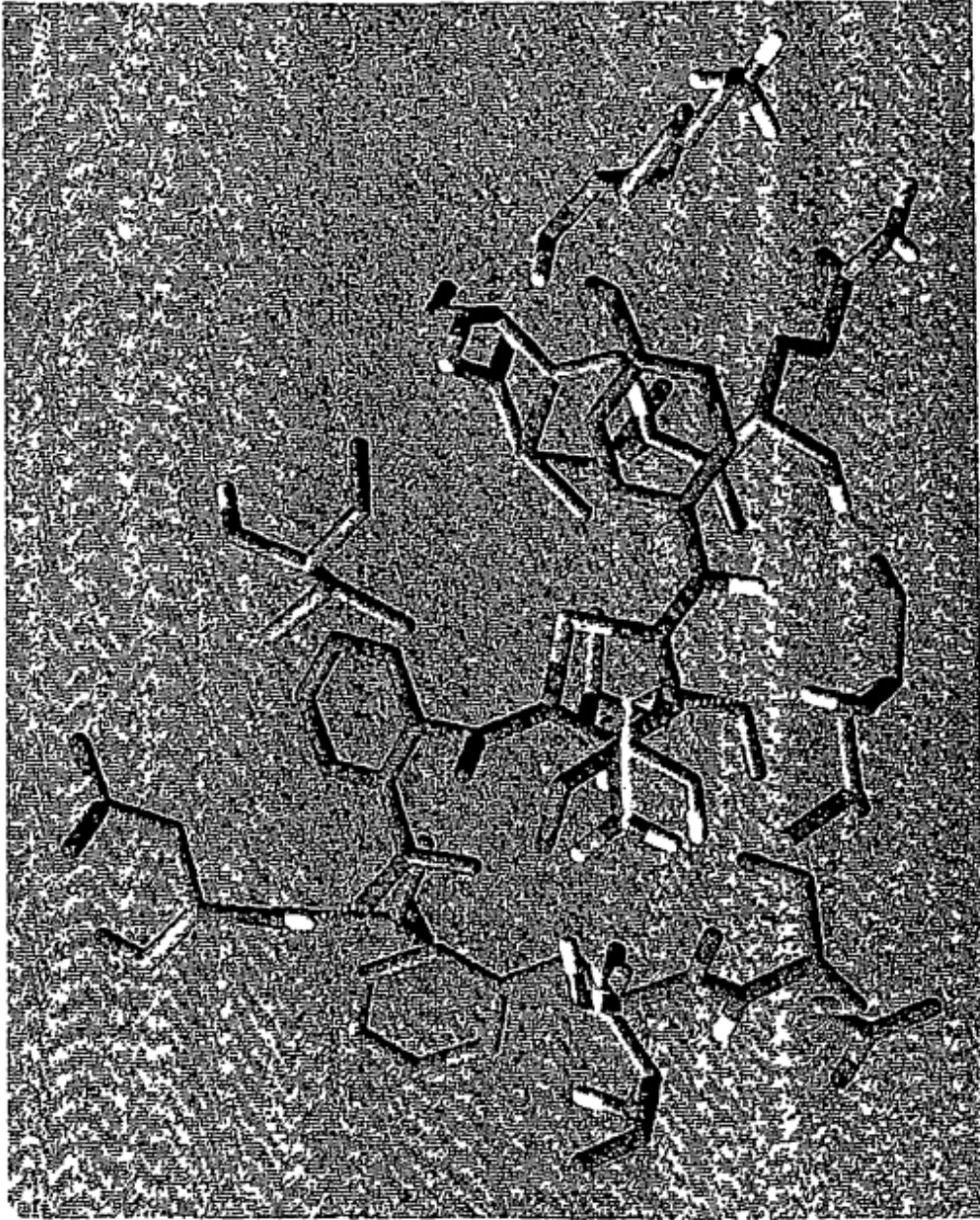


FIG. 7

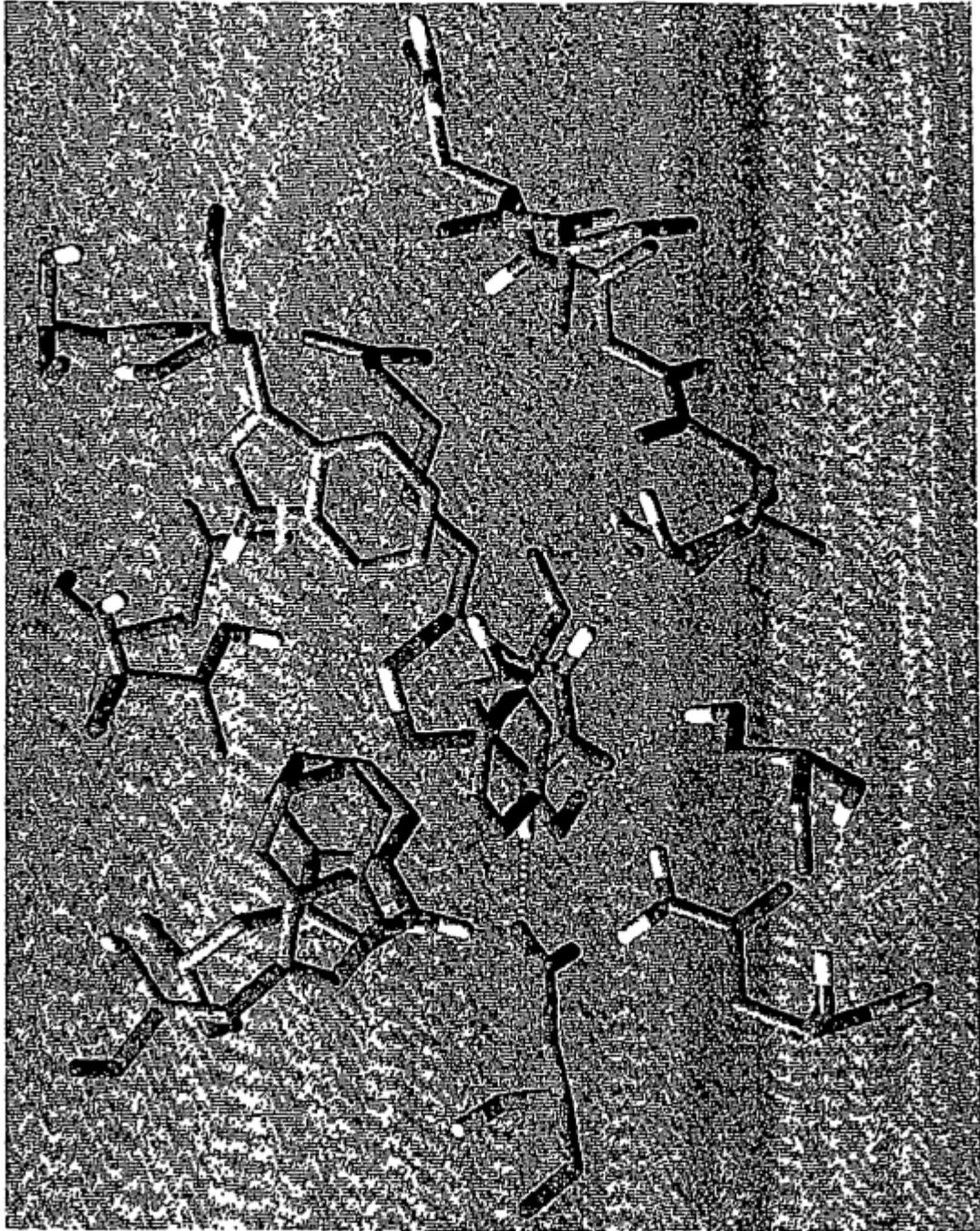


FIG. 8

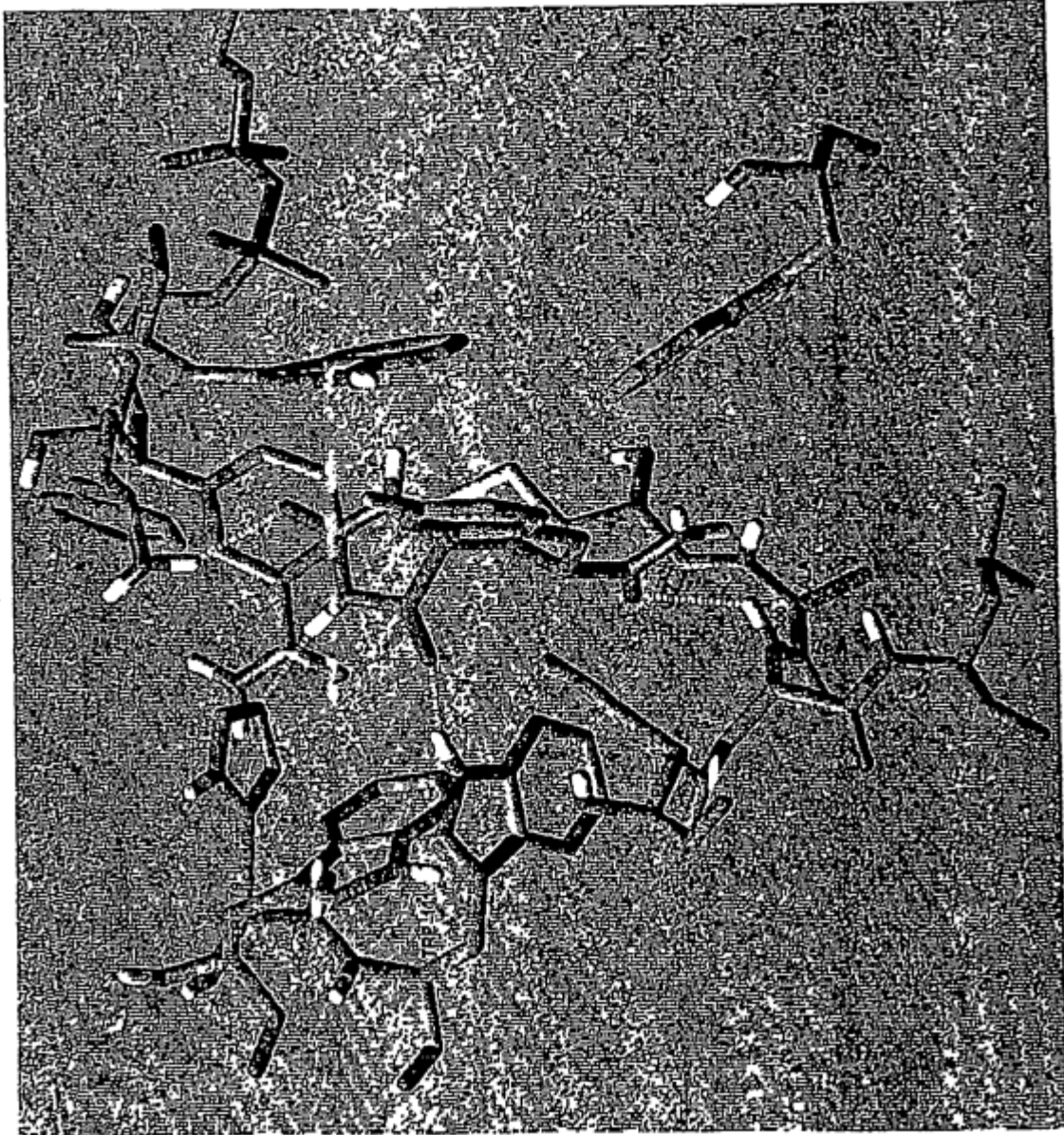


FIG. 9

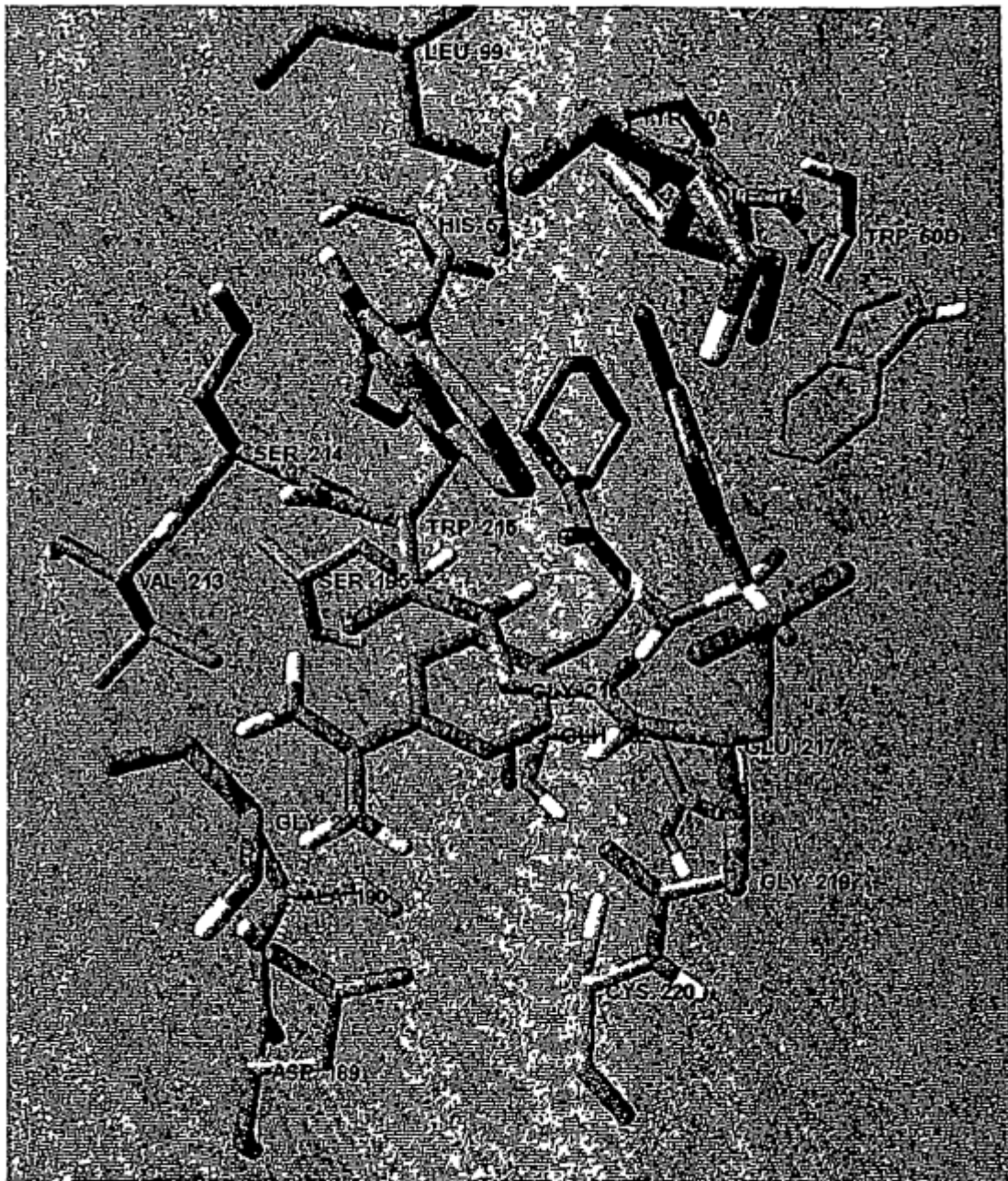


FIG. 10

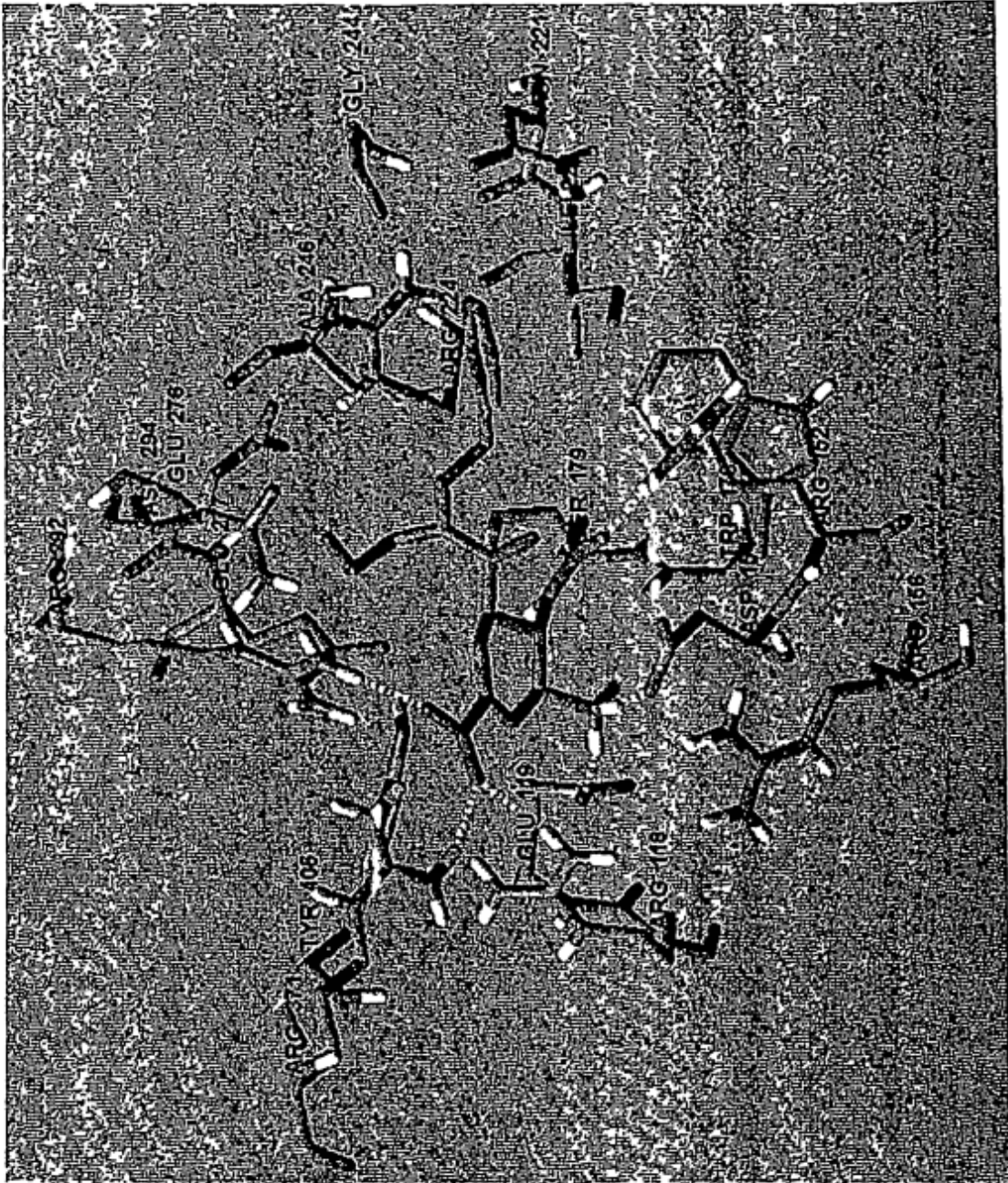


FIG. 11