

Métodos computacionales de modelado molecular y diseño de fármacos

FEDERICO GAGO BADENAS

Departamento de Fisiología y Farmacología,

Universidad de Alcalá de Henares, Madrid.

1.0 INTRODUCCIÓN.

Cualquier acción farmacológica tiene su inicio en la formación de un complejo entre la molécula de fármaco y su sitio receptor en una macromolécula biológica. Por lo tanto, la especificidad de la respuesta a un fármaco dado viene determinada en gran medida por la capacidad de los distintos receptores celulares para reconocerlo como agonista o antagonista y evocar o no una respuesta. La formación de complejos en los que participan moléculas biológicas no es un fenómeno exclusivo de la Farmacología, sino que se contempla en otras ramas de la Biología, como la Bioquímica (*e.g.* complejos de Michaelis entre enzimas y sustratos) o la Inmunología (*e.g.* complejos antígeno-anticuerpo), y de la propia Química, donde las interacciones entre moléculas "huéspedes" (*guests*) y "hospedadoras" (*hosts*) han abierto nuevas perspectivas en la denominada Química Supramolecular (Cram, 1988). Lo que confiere a la Farmacología un carácter propio es el efecto que surge como consecuencia de esa unión, y su posible manipulación con fines terapéuticos. Con la progresiva identificación de los mediadores celulares implicados en estas respuestas, la creciente caracterización de los diferentes receptores celulares involucrados, y la mejor comprensión de los mecanismos transductores y amplificadores de la señal inicial, la Farmacología moderna persigue profundizar en el conocimiento de los pasos intermedios que separan la formación del complejo fármaco-receptor de la respuesta fisiológica o bioquímica observada.

El receptor debe ser contemplado, pues, como un sitio discriminador capaz de distinguir entre posibles ligandos y moléculas que no poseen afinidad de unión. Pero además, hay que considerarlo como el primer responsable de la serie de acontecimientos que pueden traducir esa interacción en una respuesta celular. Desde el punto de vista de los

ligandos, esto quiere decir que su fijación al receptor no va a conducir invariablemente a la producción de una respuesta máxima, sino que existirá, en función de su eficacia, una gradación de respuestas, que servirá para su clasificación como agonistas, agonistas parciales o antagonistas.

La síntesis de nuevos compuestos con capacidad para interaccionar con receptores específicos es el objetivo prioritario de la llamada Química Farmacéutica o Química Médica. Hasta hace relativamente poco tiempo, estas síntesis se realizaban al azar o, en el mejor de los casos, estaban guiadas por el estudio más o menos cuantitativo del efecto que diversas modificaciones introducidas en la estructura de distintos fármacos o ligandos endógenos conocidos ejercían sobre una determinada acción farmacológica. En gran medida, esto era la consecuencia lógica del desconocimiento existente sobre la estructura de los receptores.

En los últimos años el panorama ha cambiado sustancialmente gracias a los avances en muchas y variadas disciplinas, que han hecho posible la identificación de numerosas macromoléculas diana y el conocimiento de su secuencia de nucleótidos o aminoácidos, e incluso, en algunos casos, la elucidación a nivel atómico de su estructura y la de sus complejos con inhibidores. Todo un arsenal de nuevas metodologías, con un importante componente matemático y computacional, hacen uso de esta información para crear modelos tridimensionales de receptores y ligandos, estudiar sus preferencias conformacionales, dilucidar la naturaleza y magnitud de las fuerzas interatómicas que gobiernan su interacción, y analizar el comportamiento dinámico de cada molécula por separado y de sus respectivos complejos. Estos procedimientos ayudan a comprender mejor el comportamiento de estos sistemas a nivel submolecular, permiten establecer comparaciones entre teoría y datos experimentales, e incluso permiten realizar predicciones cuantitativas, por lo que constituyen herramientas muy poderosas para diseñar nuevas moléculas con afinidad por un determinado receptor.

2.0 RELACIONES ENTRE ESTRUCTURA QUÍMICA Y ACTIVIDAD BIOLÓGICA.

Muchos de los fármacos hoy disponibles en la industria farmacéutica fueron caracterizados en su día mediante técnicas de cribado (*screening*) convencionales, consistentes en evaluar en una batería lo más amplia posible de ensayos biológicos el mayor número posible de sustancias, tanto de origen natural como sintético, elegidas más

o menos al azar. Con este procedimiento se consiguen identificar nuevos "cabezas de serie", o moléculas prototipo pertenecientes a una clase estructural determinada y con potencial en una área terapéutica concreta. Modificaciones químicas subsiguientes tienden a producir "análogos" de esas estructuras con una mayor actividad o una menor incidencia de efectos colaterales. Este método de descubrimiento de nuevos agentes con actividad biológica es interesante desde el punto de vista de que puede convertir a nuevas clases estructurales de compuestos en fármacos potenciales pero, al estar basado fundamentalmente en técnicas de ensayo y error, consume mucho tiempo y requiere grandes recursos económicos. El porcentaje de éxitos se ha estimado inferior a 1 por cada 10.000 compuestos sintetizados (Sheridan, 1987).

Los primeros intentos dirigidos a incrementar la probabilidad de sintetizar un análogo activo o de encontrar un nuevo cabeza de serie se basaron en encontrar correlaciones entre la estructura química de una serie de compuestos y su actividad biológica. De ahí surgieron las famosas siglas QSAR, acrónimo de *Quantitative Structure-Activity Relationships*, que es hoy día una palabra de uso corriente tanto en el proceso de diseño de nuevos fármacos como en la racionalización de las propiedades farmacológicas de una serie de sustancias. En los últimos 10 años hemos asistido a una auténtica revolución en el desarrollo de los métodos de QSAR, y en la actualidad todas las compañías farmacéuticas y centros de investigación importantes hacen uso de alguna de estas tecnologías para optimizar la rentabilidad de sus síntesis y mejorar sus expectativas de dar con una nueva sustancia que pueda ser explotada comercialmente. La creciente sensibilidad social ante nuevos procesos patológicos, como el síndrome de inmunodeficiencia adquirida o la enfermedad de Alzheimer, la prevalencia de altas tasas de mortalidad por distintas enfermedades, como las cardiovasculares y neoplásicas, y el descubrimiento de nuevas dianas farmacológicas, como la NO sintetasa o ciertas proteinasas virales, son algunos de los factores que proporcionan un ímpetus adicional para que la búsqueda de nuevos agentes terapéuticamente eficaces sea más racional.

Los métodos que relacionan la estructura química con la actividad biológica pueden dividirse en dos grandes categorías (Sheridan, 1987): métodos topológicos/estadísticos y métodos de modelado molecular. En la *aproximación topológica* sólo se tiene en cuenta la estructura química "plana" de la molécula y se utilizan técnicas estadísticas o de reconocimiento de patrones para encontrar las QSAR. En los *métodos de modelado*, que son el objeto de esta pequeña revisión, se consideran las propiedades de las moléculas en tres dimensiones y son importantes, entre otros, el análisis

conformacional, la mecánica cuántica, los campos de fuerzas, la termodinámica estadística, y los gráficos moleculares interactivos. Estos últimos permiten la representación y manipulación de las moléculas en tres dimensiones, lo que proporciona una información espacial que es esencial para comparar moléculas y para estudiar la interacción entre ligandos y receptores macromoleculares (Langridge, 1981). El término *modelado molecular* no es exclusivo de los estudios encaminados al diseño de nuevos fármacos, sino que es aplicable en otras áreas de investigación, como la ingeniería de proteínas o la química de polímeros. Las siglas CAMD (*Computer-Aided Molecular Design*) engloban a un número considerable de procedimientos, basados en el uso de ordenadores, encaminados a relacionar actividad con estructura molecular. Como su propio nombre indica, su fin último es utilizar estas relaciones para predecir compuestos con un determinado perfil de actividades (Hopfinger, 1985).

Una vez conseguido un ligando con alta afinidad por un determinado sitio receptor en una macromolécula biológica de interés, queda un largo camino hasta que el fármaco potencial pueda integrarse como medicamento en el arsenal terapéutico. Unas características farmacocinéticas inadecuadas, la aparición de efectos secundarios inaceptables, o la biotransformación a un metabolito tóxico, son algunos de los factores que pueden hacer que el compuesto, en principio prometedor, vuelva a los laboratorios de investigación en un intento de optimización.

3.0 MÉTODOS DEDUCTIVOS. HIPÓTESIS DEL FARMACÓFORO.

Es un hecho bien conocido que ante un receptor de estructura desconocida (el problema más frecuente hasta hace pocos años), la variación sistemática de la estructura química de sus ligandos lleva rápidamente a la conclusión de que algunas partes de la molécula son críticas para la actividad, mientras que otras pueden modificarse y su alteración sólo conlleva pequeñas variaciones de afinidad. Estas diferencias, que se manifiestan con facilidad en los resultados de los ensayos de fijación de ligandos marcados (*binding*) o de otras pruebas farmacológicas, llevaron a postular el concepto de *farmacóforo*, entendido como el conjunto de grupos químicos que todas las moléculas activas tienen en común y que son esenciales para el reconocimiento por parte del receptor. Esta formulación es semejante a la analogía entre llave y cerradura que sirvió inicialmente para comprender la complementariedad entre los sitios activos de las enzimas y sus sustratos.

La aproximación clásica de Hansch y colaboradores a las QSAR (Hansch, 1979) puso de manifiesto que un marco de referencia común basado en una serie congenérica ofrece una buena base para la interpretación racional. A principios de los 80, Hopfinger demostró que la calidad de las QSAR derivadas utilizando únicamente características físico-químicas y subestructurales mejoraba sustancialmente si se consideraba de forma sistemática la forma molecular. La base para comparar formas moleculares era la determinación del volumen común de solapamiento estérico entre pares de moléculas en función de la geometría intermolecular (Hopfinger, 1980). La diferencia de actividad entre dos moléculas cualesquiera se asumía función de las diferencias en medidas de las correspondientes propiedades moleculares, entre las que se encontraban los descriptores de forma entre pares de moléculas. Para tener en cuenta la flexibilidad conformacional se incluían en la regresión todos los conformeros estables calculados por mecánica molecular (Buttershell, 1981).

La limitación inherente a los procedimientos de minimización de la energía es que encuentran una solución que no es única y que depende de la formulación del problema. Para determinar el conjunto de todos los posibles patrones tridimensionales que pueden presentar los grupos farmacofóricos se hace imprescindible la exploración sistemática de las conformaciones asequibles a un compuesto activo (*cf.* sección 8). La decisión respecto al umbral energético elegido para rechazar una determinada conformación es también arbitraria, y además hay que tener en cuenta que la superficie energética del fármaco aislado se puede perturbar por la interacción con el receptor. Para Hopfinger, el conformero específico finalmente seleccionado era el resultado de optimizar el ajuste de regresión entre la actividad y los descriptores, uno de los cuales era precisamente el volumen común de solapamiento estérico (Hopfinger, 1983). Los métodos que calculan la semejanza entre moléculas están ya firmemente establecidos en CAMD (Burt, 1990; Good, 1993) y son de utilidad tanto para generar parámetros útiles en QSAR como para optimizar la superposición entre estructuras (Manaut, 1991; Sanz, 1993).

Buscando el equivalente topográfico o tridimensional de las QSAR, Marshall y colaboradores desarrollaron la conocida como *aproximación de los análogos activos* ('*active analog approach*'), en la que el farmacóforo proporciona un mínimo marco geométrico de referencia, o un conjunto de reglas de orientación, para la comparación de distintas moléculas no necesariamente congenéricas (Marshall, 1985). La hipótesis de partida es la complementariedad estructural entre el receptor y los compuestos que se unen a él, asumiendo además que el sitio receptor permanece constante y que fármacos

estructuralmente diferentes se unen en conformaciones que presentan un patrón estérico y electrónico semejante, el *farmacóforo*. La principal dificultad en la práctica radica en que la libertad conformacional inherente a la mayor parte de los fármacos les hace capaces de presentar multitud de patrones tridimensionales frente a un receptor dado. Cada una de las conformaciones energéticamente accesibles delimita una disposición tridimensional determinada de los grupos funcionales candidatos a interaccionar con el receptor (un heteroátomo, el centroide de un anillo aromático, un par de electrones sin compartir, etc).

Si la hipótesis del farmacóforo es cierta, cada análogo activo debería presentar un patrón que tendría que aparecer en el conjunto de posibles patrones determinados para cada compuesto. Un punto de referencia inicial lo constituyen las conformaciones de mínima energía. Si no se encuentra una disposición común de los grupos elegidos como esenciales en los conformeros de baja energía de cada molécula, habrá que revisar la selección inicial de grupos. Si se encuentra más de una disposición posible de los mismos grupos, habrá que introducir más restricciones. Sin embargo, no resulta realista asumir que los ligandos siempre se unen al receptor en sus conformaciones de mínima energía ya que la energía libre de asociación generalmente supera la energía necesaria para que el ligando sufra un cambio conformacional. Es necesario por tanto considerar todas las conformaciones que posean una energía hasta unas pocas kilocalorías/mol por encima del mínimo global. Adicionalmente hay que examinar el volumen del sitio de unión que está realmente disponible para ser ocupado por los ligandos. El volumen mínimo lo proporciona la unión de los volúmenes de todas las moléculas activas en aquellas conformaciones que maximizan la superposición de grupos farmacofóricos equivalentes (Figura 1).

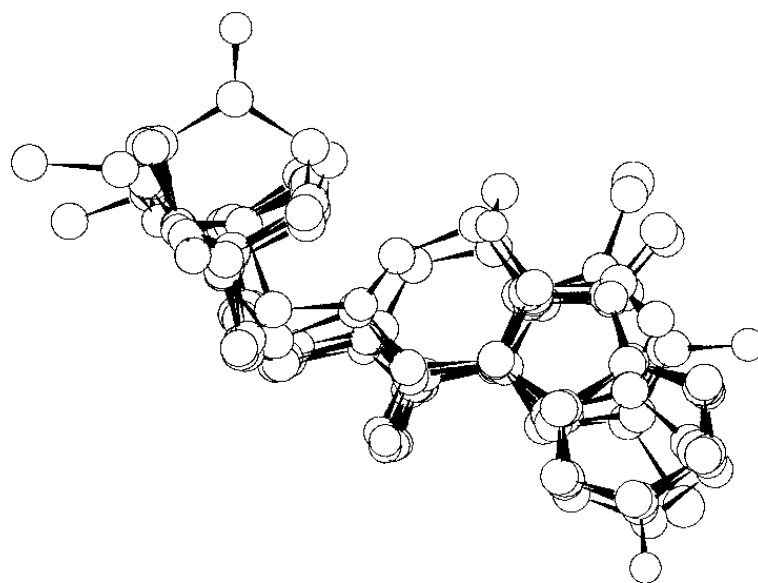


Figura 1. Superposición de 9 moléculas estructuralmente diferentes con afinidad por el receptor 5-HT₃ (Hibert, 1990). La adición del volumen esencial para cada fármaco activo cuando se encuentra unido adecuadamente determina el mínimo espacio disponible. Una molécula que posea el farmacóforo, y que incluso pueda adoptar la geometría óptima para la unión, puede resultar inactiva si se extiende más allá de los límites impuestos por este volumen mínimo del sitio activo.

Una vez los datos experimentales demuestran la validez de la hipótesis del farmacóforo, se deben examinar, para aumentar la consistencia del modelo, compuestos que posean los grupos funcionales requeridos pero que muestren poca o ninguna actividad. En el supuesto de que alguno de ellos sea, no obstante, capaz de asumir una conformación energéticamente razonable en la cual los grupos farmacofóricos se encuentran correctamente alineados con los de los compuestos activos, una explicación plausible para su inactividad es que esa molécula en esa conformación requiere un volumen adicional que está de hecho ocupado por el receptor. Este tipo de moléculas son útiles para ayudar a determinar la localización del mínimo espacio ocupado por el receptor en relación con el farmacóforo. Si el conjunto de moléculas estudiadas es suficientemente grande, se pueden llegar a determinar mapas de volumen para diversos receptores, cuya comparación puede permitir optimizar la actividad de un compuesto con respecto a otro. De este modo, esta técnica se convierte en un medio eficaz de caracterizar la topografía de un sitio receptor de naturaleza en principio desconocida (*receptor mapping*) al permitir determinar tanto el

volumen asequible al farmacóforo como el volumen adyacente que debe ser excluido. Más aún, no sólo se obtiene información sobre la forma del sitio de unión sino también sobre la posible naturaleza de los grupos funcionales presentes en el receptor, responsables de la afinidad. Sobre esta base se ha llegado a proponer, por ejemplo, que existe un farmacóforo común para 14 clases diferentes de fármacos que actúan sobre el sistema nervioso central (Lloyd, 1986). El responsable de este modo de unión común sería un receptor precursor común de aminas biógenas. Los diferentes receptores distinguirían unas moléculas de otras gracias a diferencias en sitios de unión accesorios. El reciente modelado de un buen número de receptores de membrana que funcionan acoplados a proteínas G y responden a diferentes neurotransmisores (Trumpf-Kallmeyer, 1992) proporciona una base estructural sólida a esta hipótesis.

Una vez delineado, el farmacóforo candidato (o su modelo simplificado de receptor con grupos funcionales complementarios en las posiciones adecuadas) puede guiar en la interpretación de la actividad de compuestos químicos diversos y en el desarrollo de nuevas clases de agentes terapéuticos. Este concepto sirvió, por ejemplo, para explicar la actividad promotora de tumores de los esteres de forbol, teleocidina B y aplisiatoxina, en base a su semejanza con el activador natural de la proteína quinasa C, el 1,2-diacilglicerol (Wender, 1986), o para la definición geométrica de un modelo para el sitio activo de la enzima convertidora de angiotensina (ECA) en el que se ajustaban 28 clases distintas de inhibidores (Mayer, 1987). Estas técnicas han servido asimismo para la caracterización topográfica del receptor 5-HT_{1A} de serotonina, que condujo a la síntesis de un compuesto con una selectividad mejorada por este receptor frente al adrenoceptor α_1 (Hibert, 1988), así como para definir el farmacóforo de una serie de antagonistas anti-5HT₃ (Hibert, 1990).

Una limitación de esta aproximación es la simplificación que supone la superposición de grupos funcionales de los ligandos cuando en realidad son posibles modos de unión alternativos, como se ha puesto de manifiesto claramente mediante el análisis de estructuras cristalinas de complejos proteína-ligando (Badger, 1988) o ADN-ligando (Quintana, 1991), o incluso unión a diferentes sitios de un mismo receptor (Marullo, 1990). Otra limitación es que, al ignorar las afinidades relativas de los diferentes compuestos, se pierde información y no se puede cuantificar la bondad de la interacción con el supuesto sitio receptor.

Una forma de paliar esta última limitación consiste en utilizar la metodología de geometría de distancias y los conceptos de QSAR para generar un sitio hipotético con unas

propiedades tales que su afinidad calculada para series de análogos reproduzcan las afinidades experimentales. De este modo, el sitio modelado puede ayudar en la selección de compuestos a sintetizar e investigar y permite su continuo refinamiento a medida que se van obteniendo nuevos datos.

3.1. Geometría de distancias.

La serie de ligandos se puede representar como una colección de puntos en el espacio que corresponden a átomos o pequeños grupos de átomos, y la flexibilidad conformacional se puede tratar como límites máximos y mínimos sobre las distancias entre todos los pares de puntos que constituyen el ligando (Crippen, 1977). Para manejar la información contenida en este tipo de matrices se hace uso de la metodología de geometría de distancias, que Crippen aplicó por vez primera al estudio de la interacción ligando-receptor (Crippen, 1979, 1980). Comparando las matrices de distancias de un número suficiente de ligandos de interés (que deben pertenecer a clases estructurales diferentes) es posible deducir sus características estructurales comunes, con respecto a las cuales se pueden ir posicionando los sustituyentes y los sitios complementarios del receptor. Suponiendo que la energía total de unión del ligando a su receptor es igual a la suma de las interacciones individuales entre sitios del ligando y sitios del receptor, y conocida la energía libre de unión a partir de la constante de disociación del complejo ligando-receptor determinada experimentalmente, el método permite calcular las contribuciones energéticas individuales de cualquier interacción entre estos sitios. Las limitaciones más importantes de este tipo de cálculos es que no permiten cambios conformacionales en el receptor y no tienen en cuenta la energía conformacional del ligando, que se asume es pequeña con relación a la energía libre de unión.

Además de incorporar en los cálculos la flexibilidad conformacional del ligando, el método permite al investigador, en condiciones ideales, proponer nuevos ligandos con mayor afinidad sobre la base de los parámetros geométricos y energéticos del modelo deducido del sitio receptor. El sitio de unión en la macromolécula se propone en base a un número de sitios puntuales que ocupan posiciones fijas en el espacio unos respecto a otros y pueden estar "vacíos" o "llenos". Un sitio puntual vacío es un lugar disponible que se encuentra en una posición donde se puede situar el ligando cuando tiene lugar la unión. Un sitio lleno, por el contrario, indica la posición de algún grupo que produce bloqueo estérico, por lo que ningún punto del ligando puede coincidir con él durante la unión. Los

posibles modos de unión de los ligandos vienen dictados por aquellos puntos del ligando que ocupan sitios puntuales vacíos del receptor. La suma de las contribuciones de todos los contactos entre cada punto del ligando y puntos del receptor proporciona la energía libre calculada. El mayor número de puntos comunes a todas las moléculas de ligando que tienen los mismos tipos de sitios y distancias entre ellos constituye el grupo base o farmacóforo, que se asume se une siempre al receptor en la misma orientación. A continuación se determina el conjunto más pequeño de grupos sustituyentes necesarios para considerar los puntos restantes de todos los ligandos, se evalúan los límites de distancia de los sitios puntuales y se deducen las coordenadas (Donné-Op den Kelder, 1988).

3.2. QSAR tridimensional. Análisis comparativo de campos moleculares.

En la aproximación de los análogos activos presentada anteriormente, la descripción de la actividad de los compuestos se limita a una elección binaria: activos o inactivos. En la práctica, sin embargo, cuando se miden las respuestas mediadas por un receptor o su afinidad de unión ante distintos ligandos, se encuentra una amplia dispersión en los resultados. Estas diferencias reflejan las sutiles variaciones que tienen lugar en la interacción y pueden ser muy clarificadoras si se consiguen interpretar de forma consistente. Tradicionalmente éstos han sido los dominios de los estudios de QSAR, que ofrecen una base para la interpretación racional pero eluden la cuestión conformacional al centrarse en series congenéricas en las que la relación tridimensional de los distintos fragmentos moleculares se asume que es semejante debido al esqueleto químico común (Hansch, 1986). La necesidad de pasar de esta aproximación esencialmente topológica a un equivalente topográfico aplicable a compuestos con estructuras no necesariamente relacionadas y que tuviera además en cuenta explícitamente la variable conformacional obligó a desarrollar un marco conceptual y computacional diferente (Motoc, 1986; Labanowski, 1986).

La aproximación de los análogos activos sienta las bases para comparar series no congenéricas de compuestos, y hace posible asociar cambios en potencia con la localización de diversos fragmentos moleculares en distintas regiones del espacio del receptor. Otro intento de describir las propiedades de unión de una serie de compuestos de forma cuantitativa, y un paso adelante para la predicción de nuevos compuestos biológicamente activos sobre la base de compuestos previamente sintetizados y ensayados,

es el análisis comparativo de campos moleculares (*CoMFA* = *Comparative Molecular Field Analysis*), desarrollado por Cramer y colaboradores en 1988. La premisa básica en este método es que "el adecuado muestreo de los campos estérico y electrostático alrededor de un conjunto de moléculas de ligando (o de fármaco) puede proporcionar toda la información necesaria para comprender sus propiedades biológicas" (Cramer, 1988). La lógica subyacente radica en que las interacciones que dan lugar a la producción de un efecto biológico suelen ser no-covalentes y en que una gran variedad de propiedades moleculares se pueden explicar de forma precisa mediante campos de fuerzas de mecánica molecular que sólo consideran fuerzas estéricas y electrostáticas.

Para llevar a cabo el muestreo, se calculan las energías de interacción tanto estérica (van der Waals 6-12) como electrostática (Coulomb, con una constante dieléctrica inversamente proporcional a la distancia) entre cada molécula de interés y una "sonda atómica" adecuada (*e.g.* con las propiedades de un carbono sp^3 y carga +1.0) situada en puntos regularmente espaciados (*e.g.* 2 Å) de una malla tridimensional que rodea a todos los compuestos sometidos a examen. Para ello es necesario que todos los ligandos se alineen previamente de tal modo que, en el mejor de los casos, el alineamiento sea equivalente a la orientación espacial relativa de las diversas moléculas de fármaco en el sitio receptor. En otras palabras, se lleva a cabo una superposición de las moléculas en conformaciones que presumiblemente se corresponden con las biológicamente activas. Las propiedades de unión de los ligandos se suelen expresar en función de sus constantes de asociación, directamente relacionadas con las energías libres de unión (*vide infra*). Como el análisis de regresión lineal está limitado por el gran número de variables existentes para el conjunto de observaciones experimentales disponible, se recurre a las nuevas técnicas estadísticas de mínimos cuadrados parciales (PLS) para extraer las variables latentes necesarias para la descripción óptima de la varianza del conjunto de datos, las cuales se someten posteriormente a un método de validación cruzada para eliminar el riesgo de correlaciones casuales.

Los pasos a seguir para un análisis de este tipo serían los siguientes (Donné-Op den Kelder, 1988):

- a) postular un conjunto de reglas de alineamiento (de forma semejante a la aproximación de los análogos activos),
- b) alinear las moléculas y establecer a su alrededor una malla tridimensional como una caja de puntos en el espacio potencial del receptor,
- c) calcular el campo que cada molécula ejercería sobre un átomo sonda situado sobre cada

punto de la malla (normalmente se guardan por separado los efectos estéricos y los electrónicos para facilitar su interpretación),

d) utilizar estadística de PLS para determinar una expresión lineal ("QSAR tridimensional"), que podría consistir simplemente en un conjunto mínimo de puntos de la malla necesarios para distinguir a los compuestos sometidos a examen de acuerdo con sus actividades determinadas experimentalmente.

e) validar de forma cruzada el modelo PLS (*cross-validation*), es decir, comprobar su valor predictivo eliminando observaciones sucesivamente, rederivando el modelo, y "prediciendo" las observaciones eliminadas.

f) ajustar ligeramente el alineamiento del compuesto o compuestos peor predichos y repetir los pasos anteriores hasta que se encuentre un alineamiento mutuo que produzca una QSAR-3D con alto valor predictivo (alto valor de *cross-validated* r^2).

Las QSAR producidas por un CoMFA, con sus cientos o miles de términos, se pueden visualizar gráficamente de forma muy simple, representando los mapas de contorno para los valores especificados de los coeficientes, que además pueden colorearse diferencialmente de acuerdo con la dirección y la magnitud de la interacción. Los poliedros coloreados resultantes en cada mapa mostrarán aquellas regiones donde las diferencias de actividad observadas se encuentran más relacionadas con cambios en los campos estéricos o electrostáticos. Estos diagramas de contorno se pueden interpretar en términos de estructura del receptor en aquellas ocasiones en que la estructura del receptor es conocida (Marshall, 1988).

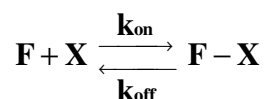
La variable que resulta más crucial en CoMFA es el posicionamiento de las moléculas dentro de la malla tridimensional (Cramer, 1988; Klebe, 1993). En ausencia de información experimental sobre el modo de unión de la serie de moléculas sometidas a examen, se pueden intentar superponer los grupos farmacofóricos equivalentes minimizando simultáneamente las distancias entre los átomos que los definen en las distintas moléculas y la energía potencial de cada molécula individual ("ajuste múltiple flexible"). Los resultados de este procedimiento van a estar sesgados por la selección que se haya hecho del farmacóforo y van a ser dependientes de las geometrías de partida. Otro procedimiento, tendente a incrementar la semejanza de campos dentro de la serie estudiada ("ajuste de campos"), minimiza la diferencia de mínimos cuadrados en la suma de las energías de interacción estérica y electrostática, promediadas para todos los puntos de la malla, entre una determinada molécula del conjunto y el promedio de todas las moléculas del conjunto (Cramer, 1988). Para ello las moléculas se tratan como cuerpos rígidos que

sólo pueden trasladarse o rotarse en las tres direcciones del espacio. Este segundo procedimiento se suele utilizar para modular el ajuste puramente estérico del primero, con lo que se tienen mejor en cuenta las diferencias en las propiedades electrostáticas de las moléculas consideradas.

Otro aspecto fundamental a considerar es que las energías de interacción calculadas en CoMFA mediante mecánica molecular no incluyen factores con una base entrópica, como la hidrofobicidad. Como la constante de asociación está relacionada con la energía libre de unión, a lo más que se puede aspirar es a correlacionar la energía de interacción calculada con el componente entálpico de la energía libre. En la práctica esto significa que sólo se podrán predecir adecuadamente entalpías de unión, y no energías libres o constantes de unión. Es de suponer que moléculas que presenten diferentes grados de flexibilidad conformacional, es decir que exhiban un número diferente de conformaciones accesibles en el estado no unido, mostrarán diferentes contribuciones entrópicas a la energía libre de unión una vez que se encuentren inmovilizadas conformacionalmente en su sitio de unión. El no considerar estos efectos entrópicos adecuadamente en CoMFA puede traducirse en un peor alineamiento y en la caracterización de una conformación que no es exactamente la conformación "activa" (Klebe, 1993).

4.0 EVALUACIÓN EMPÍRICA DE LAS INTERACCIONES ENTRE GRUPOS FUNCIONALES DE LIGANDOS Y RECEPTORES.

La afinidad de un fármaco F por un receptor X en la reacción de asociación simple



viene dada por

$$K = \exp(-\Delta G^0/RT) \quad [1]$$

donde $K = [\text{FX}]/[\text{F}][\text{X}]$ es la constante de asociación en el equilibrio, ΔG^0 es el cambio en energía libre de Gibbs, R es la constante de los gases (producto de la constante de Boltzmann por el número de Avogadro) y T es la temperatura en Kelvin. La energía libre de Gibbs es una medida adecuada de la fuerza directriz para reacciones a temperatura y presión constantes. La especificidad de la interacción entre F y X viene dictada por la diferencia entre los valores de K correspondientes a la unión de ligandos específicos y no-

específicos. Las constantes de unión más altas de los ligandos específicos surgen de fuerzas de atracción más fuertes entre ambas moléculas, la mayor parte de las cuales son de corto alcance (distancias menores de 2-4 Å): fuerzas de dispersión de van der Waals-London, efecto hidrofóbico y enlace de hidrógeno. De estos tipos de fuerzas, los dos primeros pueden considerarse isotrópicos, pero los enlaces de hidrógeno son estrictamente direccionales y requieren la orientación de los grupos participantes. Las fuerzas electrostáticas son de largo alcance, pero en la realidad se ven limitadas por la alta constante dieléctrica del agua. Por ello, el enterramiento de grupos cargados en una interfaz ligando-receptor (o antígeno-anticuerpo, o enzima-sustrato) conlleva un aumento del campo electrostático efectivo, pero a costa de la desolvatación energéticamente desfavorable de los grupos cargados. Por esta razón, el complejo estable sólo se formará entre ambas moléculas si, además de presentar complementariedad de formas, su distribución de cargas superficiales es tal que la interacción de cargas opuestas proporciona una atracción de Coulomb suficiente para compensar la pérdida en energía libre de solvatación y así estabilizar el complejo (Novotny, 1992).

Los sitios de unión en macromoléculas son prácticamente siempre cavidades o surcos. La preferencia por superficies cóncavas se puede explicar por tres motivos fundamentales: i) la difusión para salir de la cavidad es significativamente más lenta que cuando se trata de una superficie plana o es a través del disolvente, ii) en las cavidades, los campos electrostáticos se encuentran aumentados o focalizados, y iii) las superficies cóncavas son más hidrofóbicas que su equivalente plano o convexo porque la hidrofobicidad de una superficie es función de su curvatura con respecto al tamaño y curvatura de una molécula de agua.

Para estimar las fuerzas aproximadas de las interacciones fármaco-receptor, la alternativa más simple (Albert, 1979) es la de utilizar los valores estándar de las entalpías de formación de los diferentes tipos de enlace implicados (iónico, de hidrógeno, de van der Waals, etc). Esta aproximación, sin embargo, no incluye ningún componente entrópico y presenta además incertidumbres en cuanto a cuáles de los posibles enlaces deben ser incluidos en la interacción y cuál debe ser el valor preciso de la energía asociada con cada uno de ellos. Para superar estas dificultades se han intentado estimar las fuerzas de las interacciones que implican a grupos funcionales individuales del fármaco. El "principio del ancla", desarrollado por Page en la década de los 70, se basa en encontrar pares de compuestos para los cuales la diferencia en afinidades de unión a un receptor se pueda achacar a la contribución de un único grupo funcional (Page, 1977). Para ello es necesario

disponer de muchos pares de moléculas que difieran entre sí sólo en la presencia de un grupo funcional determinado. La diferencia en energía de unión entre dos fármacos, con y sin ese grupo funcional, debe incorporar sólo aquellos factores asociados con ese grupo, al tiempo que excluye la pérdida de entropía rotacional y traslacional asociada con la molécula de fármaco (el "ancla"). Aparte de las dificultades inherentes a conseguir la serie de compuestos necesaria para un análisis de este tipo, la mayor objeción a esta metodología es asumir que el papel del grupo funcional estudiado se limita exclusivamente a proporcionar una energía de unión adicional, olvidando, entre otros factores, la posible facilitación conformacional de una forma biológicamente activa de la molécula. Andrews y colaboradores intentaron superar esta limitación utilizando datos experimentales publicados para una serie de 200 fármacos e inhibidores enzimáticos, seleccionados subjetivamente por su buena fijación a sus correspondientes sitios receptores, para hacer un cálculo estadístico de las fuerzas de los enlaces no covalentes asociados con cada grupo funcional en un entorno fármaco-receptor medio (Andrews, 1984). Como estos agentes actúan sobre receptores distintos, no se puede aplicar el principio del ancla, haciéndose necesaria la inclusión de un factor corrector, fijado en 14 kcal mol⁻¹ en condiciones estándar, para tener en cuenta la pérdida de entropía rotacional y traslacional. También se debe incluir la pérdida de entropía de rotación interna tras la unión, ya que puede asumirse que la conformación unida de un fármaco, al menos en su receptor óptimo, será relativamente fija. Este término lo derivaron estos autores empíricamente a partir de los datos experimentales, de la misma manera que las energías de unión intrínsecas para los grupos funcionales individuales, de acuerdo con la siguiente ecuación:

$$\Delta G = T\Delta S_{rt} + n_{gdl} E_{gdl} + \sum n_X E_X \quad [2]$$

donde ΔG es la energía libre de unión determinada a partir de constantes de disociación (ecuación [1]) en ensayos de fijación o de constantes de inhibición enzimáticas ($\Delta G = -RT \ln K$), $T\Delta S_{rt}$ es la pérdida global de entropía rotacional y traslacional de la molécula de fármaco unida, n_{gdl} es el número de grados internos de libertad conformacional en la molécula de fármaco (rotaciones internas posibles, excluyendo la rotación del grupo metilo y la de sistemas restringidos como el grupo amida), E_{gdl} es la energía correspondiente asociada con el cambio en entropía que resulta de la pérdida de cada grado de libertad conformacional, y E_X es la energía intrínseca de unión asociada con el grupo funcional X, del que hay n_X presentes en el fármaco. Cada valor de E_X incorporaría

así un número de términos, entre los que se incluyen (i) la entalpía de la interacción entre el grupo funcional y su correspondiente sitio de unión en el receptor, (ii) los cambios de entalpía y entropía asociados con la eliminación de agua de hidratación, tanto del grupo funcional como de su sitio de unión, y con la formación de enlaces entre las moléculas de agua desplazadas, y (iii) la entropía vibracional de baja frecuencia asociada con los nuevos enlaces formados entre el grupo funcional y los grupos del receptor. En una primera aproximación, todos estos factores pueden considerarse como propiedades del grupo funcional estudiado relativamente independientes de los grupos a los que se encuentra unido, de modo que es razonable utilizar tales potenciales de unión intrínsecos de forma aditiva para proporcionar una medida global de la interacción fármaco-receptor. Una vez codificada toda la información para los 200 compuestos, el análisis de los datos mediante un programa de regresión lineal, de acuerdo con la ecuación [2], permitió obtener una estimación media de los valores de E_x para cada unidad estructural seleccionada (Tabla I).

Tabla I. Energías intrínsecas de unión promediadas (kcal mol^{-1})

grupo	energía
Carbono tetrahédrico	0.8
Carbono trigonal	0.7
Nitrógeno protonado	11.5
Nitrógeno neutro	1.2
Grupo carboxilato	8.2
Grupo fosfato	10.0
Grupo hidroxilo	2.5
Grupo carbonilo	3.4
Oxígeno, azufre	1.1
Halógeno	1.3
Enlace de rotación libre	-0.7

Estas energías intrínsecas de unión siguen el orden esperado, siendo los grupos cargados los que dan lugar a interacciones más fuertes, seguidos por los grupos polares y, finalmente, los apolares. El signo negativo del coeficiente para el término relacionado con el número de grados de libertad rotacional interna está justificado porque la menor libertad conformacional del fármaco tras la unión resulta en una pérdida de entropía que contribuye desfavorablemente a la energía libre de unión.

Estos valores medios de energías de unión, aunque de dudoso valor predictivo, reflejan la energía de unión esperada para un fármaco medio sobre la base de sus partes constituyentes. Una desviación positiva entre la unión observada y la calculada mediante la ecuación [2] indicaría que se une mejor que la media, sugiriendo que no todas las características estructurales del ligando se están utilizando para la unión a la macromolécula y, en el caso de moléculas flexibles, que la forma unida de la molécula es una conformación de alta energía relativa. Por otra parte, una desviación negativa sería indicativa de una unión más débil que la media, de lo que se podría deducir que la mayor parte de los grupos funcionales de la molécula de fármaco están interaccionando favorablemente con el receptor y que el fármaco probablemente actúa en una conformación de baja energía, por lo que puede ser un buen cabeza de serie para el desarrollo de nuevos ligandos.

5.0 DETERMINACIÓN DE LA ESTRUCTURA TRIDIMENSIONAL DE RECEPTORES FARMACOLÓGICOS Y DE SUS COMPLEJOS CON LIGANDOS.

Gracias al trabajo que culminó en el modelo de Watson y Crick, los ácidos nucleicos, y en particular el ADN, se distinguen por ser los primeros constituyentes realmente grandes de la célula para los que se pudo describir una estructura y una conformación detalladas (Watson, 1953). Los trabajos originales sobre fibras cristalinas (Arnott, 1972) dieron paso a la obtención y análisis de cristales de secuencias cortas de ADN (*e.g.* decámeros o dodecámeros) que han mostrado al ADN como una molécula extraordinariamente flexible, cuya estructura depende de sus interacciones con otras moléculas, y cuya deformación y deformabilidad viene dictada por su secuencia particular de nucleótidos (Dickerson, 1991). El ADN es un blanco farmacológico de suma

importancia, al que se pueden unir pequeñas moléculas por mecanismos tanto intercalativos como no-intercalativos, y casi siempre con un cierto grado de selectividad de secuencia (Pullman, 1990).

La primera descripción detallada de la forma molecular de una proteína data de 1958, cuando se aplicaron técnicas de cristalografía de rayos X a la mioglobina de ballena (Kendrew, 1958). Estos estudios mostraron al grupo prostético hemo como un ligando orientado de forma precisa en una cavidad de una proteína y sirvieron para interpretar la fisiología del transporte de oxígeno sobre la base de cambios en la interacción entre este grupo y la globina. La primera enzima no apareció en la literatura hasta 1965, cuando la resolución de la lisozima procedente de la clara de huevo (Blake, 1965) vino a corregir la falsa creencia de que sólo las globinas eran susceptibles de ser estudiadas mediante rayos X. La elección de todas estas proteínas venía dictada, por supuesto, por la disponibilidad de material que podía ser purificado y cristalizado, y no por criterios que pudieran ser relevantes para el diseño de nuevos fármacos.

Los modelos moleculares rígidos utilizados inicialmente para presentar los datos de rayos X influyeron de forma decisiva sobre los argumentos esgrimidos para racionalizar las correlaciones estructura-función de las proteínas durante la década de los 60. Sin embargo, a finales de los 70 el movimiento interno de las proteínas estaba firmemente documentado, fundamentalmente por los datos suministrados por la resonancia magnética nuclear (RMN). Un resultado importante de la investigación sobre los aspectos dinámicos de las macromoléculas biológicas ha sido la contemplación de una estructura, no como un conjunto único de coordenadas atómicas, sino acompañada de sus correspondientes fluctuaciones de posiciones atómicas, energía y entropía alrededor de unos valores medios. La resolución de la estructura en disolución del tendamistat, un inhibidor de la α -amilasa, exclusivamente mediante técnicas de ^1H -RMN y geometría de distancias (Kline, 1986) marcó el comienzo de una nueva era en la determinación de estructuras tridimensionales

de proteínas y ácidos nucleicos. Hoy día, la calidad de las estructuras determinadas en disolución con las técnicas de RMN es comparable a la conseguida con técnicas de difracción de rayos X utilizando cristales, con la ventaja añadida de que muestran transiciones estructurales que pueden ser de crucial importancia para las propiedades funcionales (Wüthrich, 1990).

En Abril de 1993, el Banco de Datos de Proteínas (PDB) de Brookhaven (Upton, NY 11973, USA), principal depositario de coordenadas atómicas de macromoléculas biológicas procedentes de cristalografía de rayos X y de resonancia magnética nuclear (Bernstein, 1977), contaba ya con 982 proteínas, y disponía de aproximadamente un centenar de estructuras de ácidos nucleicos y una decena de carbohidratos. Al ritmo exponencial de crecimiento actual, se espera que para finales de 1998 existan más de 10.000 entradas en esta base de datos. Todo este caudal de información, accesible a cualquier investigador, puede ser de gran utilidad a la Farmacología y a la Química Farmacéutica. A partir de las observaciones sobre macromoléculas y sus complejos con pequeñas moléculas, se pueden conocer las características de la unión y las interacciones que mantienen al ligando en posición. Estos complejos constituyen el punto de partida para otro tipo de estudios más predictivos encaminados al diseño de nuevos ligandos (*cf.* sección 10).

La cocrystalización de enzimas con inhibidores competitivos del sustrato proporciona mucha información estructural sobre el proceso de acoplamiento, que en muchas ocasiones conlleva cambios conformacionales significativos. Más aún, los complejos de enzimas con análogos del estado de transición del sustrato han servido como base para revelar las interacciones responsables de la catálisis, algunos de cuyos componentes dinámicos se pueden deducir comparando la estructura de la enzima acomplejada con la de la enzima libre (Lolis, 1990).

6.0 VISUALIZACIÓN Y ANÁLISIS DE LOS COMPLEJOS LIGANDO-RECEPTOR. SUPERFICIES MOLECULARES.

Los métodos más tradicionales de representación gráfica de estructuras moleculares mediante ordenador visualizan a los átomos como esferas (bien puntuales o bien con un radio proporcional al radio de van der Waals) y a los enlaces químicos como segmentos que unen los centros de estas esferas. Una alternativa es representar la forma molecular como una envoltura continua en contacto con los átomos que son accesibles al disolvente. El concepto de *superficie accesible al disolvente*, desarrollado e implementado originalmente por Lee y Richards para su aplicación al estudio del plegado de las proteínas (Lee, 1971), ha sufrido algunas matizaciones y ha servido para desarrollar algunas metodologías de interés para el diseño de fármacos. En su formulación original, la superficie accesible al disolvente se definía como el área delimitada por el centro de una sonda esférica (que representa una molécula de agua y suele tener 1.4 Å de diámetro) a medida que rueda sobre la superficie de la molécula de interés, y, en consecuencia, está desplazada de la superficie de van der Waals (Figura 2). Posteriormente, Richards definió una superficie accesible al disolvente alternativa que consiste en la parte de la superficie de van der Waals de los átomos que son accesibles a la sonda esférica (*superficie de contacto*) junto con una serie de superficies cóncavas y en forma de sillín (*superficie de reentrada*) que alisan las rendijas y oquedades entre los átomos. Esta superficie es el límite del volumen del que se excluye a la sonda esférica para que no experimente solapamiento de van der Waals con los átomos (Richards, 1977). Connolly desarrolló un algoritmo numérico para situar puntos sobre esta superficie molecular accesible al disolvente (Connolly, 1983a) e inventó un algoritmo de superficie analítica para computar áreas y volúmenes moleculares (Connolly, 1983b). La visualización de estas superficies punteadas de alta resolución es, por lo general, más útil que la representación de esferas sólidas (o modelos CPK) porque (i) las superficies son transparentes, permitiendo que se vean los

enlaces e incluso las etiquetas identificativas de los átomos, (ii) la imagen puede rotarse, escalarse y seccionarse en tiempo real, permitiendo, por ejemplo, la exploración de cavidades en el interior de la proteína, y (iii) los puntos de la superficie pueden colorearse de acuerdo con distintas propiedades químicas, como la hidrofobicidad o el potencial electrostático molecular, lo que expande la idea de complementariedad topográfica a complementariedad lipofílica o electrostática, de importancia capital en las interacciones proteína-ligando, proteína-proteína, ADN-proteína y ADN-ligando (Weiner, 1982). Además de servir como lienzo sobre el que mostrar información química de distintas clases, las superficies moleculares juegan un papel fundamental en el cálculo de esa misma información. Por ejemplo, el potencial se puede evaluar en el centro de cada posición de la sonda esférica que genera un punto de la superficie (Besler, 1990) o en una serie de capas de Connolly definidas por fuera del radio de van der Waals de los átomos (Orozco, 1990).

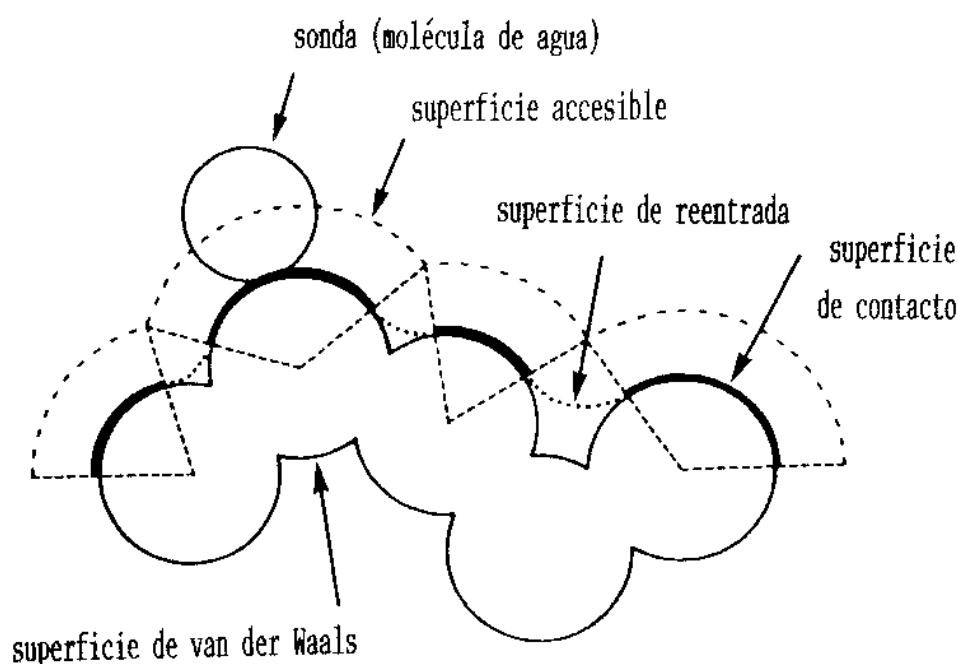


Figura 2. Definición de superficies moleculares.

Encontramos un segundo nivel de sofisticación en el análisis cuando intentamos calcular la afinidad del ligando por su sitio de unión. Para calcular la energía del complejo se suele hacer uso de la mecánica molecular, a partir de la cual podemos estimar la entalpía de unión (ΔH) sustrayendo las energías de ligando libre y receptor libre. Es evidente que estos cálculos son sólo aproximados, pues normalmente se hacen en el vacío, en tanto que las observaciones experimentales tienen lugar en disolución y se correlacionarían mejor con las energías libres de unión (ΔG). La omisión de los efectos entrópicos y de solvatación en estos cálculos teóricos refleja las limitaciones actuales de la metodología, aunque la utilización de técnicas de simulación de dinámica molecular con inclusión explícita de disolvente ya está al alcance de la mayor parte de los grupos de investigación teóricos. Cuando se estudian series congenéricas, se puede argumentar que los efectos entrópicos y de solvatación van a ser semejantes para los distintos análogos. En ocasiones la correlación encontrada entre las ΔH calculadas y las afinidades de unión es probablemente debida a la compensación entropía-entalpía observada en disolventes acuosos (Lumry, 1970; Breslauer, 1987), aunque esto no se puede asumir de forma sistemática para la interacción fármaco-receptor (Hitzemann, 1988). Entre los problemas adicionales podemos citar la inexistencia en los campos de fuerzas habitualmente utilizados para macromoléculas de parámetros para las nuevas estructuras (lo que requiere un estudio teórico intensivo), la necesidad de aplicar con rigor la mecánica cuántica para reproducir adecuadamente los efectos electrostáticos, las incertidumbres sobre la constante dieléctrica del medio que apantalla la intensidad de la interacción, la polarizabilidad de los átomos que intervienen en la formación del complejo, la posibilidad de que exista más de un modo de unión para compuestos con estructuras semejantes, la localización del mínimo global, etc.

Uno de los primeros sistemas que sirvieron para introducir la mecánica molecular

en el estudio de complejos modelados fue el de la unión de las hormonas tiroideas T₃ y T₄ a la prealbúmina (Oatley, 1984), aprovechando el hecho de que se habían determinado las estructuras de rayos X de los complejos. Así, fue posible correlacionar las afinidades relativas de T₃ y T₄ con las energías de unión calculadas y racionalizar de forma cualitativa la mayor afinidad de esta última por la proteína, una vez introducido un término corrector para tener en cuenta las diferencias de solvatación entre las dos moléculas. Los trabajos pioneros asumían que tanto el ligando como el receptor permanecían rígidos, limitando por tanto la exploración a las seis variables de traslación y orientación relativas de uno respecto a otro. La adición de flexibilidad interna al sistema permite el ajuste inducido y mejora la complementariedad, pero a costa de aumentar la complejidad y, paralelamente, la demanda computacional, aunque esto no supone un problema insalvable con la creciente capacidad de cálculo de los modernos ordenadores.

Estos procedimientos, aplicados a series con un mayor número de análogos, permiten evaluar sobre una base más firme las características de la unión y extraer conclusiones que pueden ser de utilidad para el diseño de nuevos ligandos (Menziani, 1989; Gago, 1989).

7.0 MODELADO DE LIGANDOS Y DE RECEPTORES FARMACOLÓGICOS.

Las estructuras tridimensionales de los ligandos, en caso de no disponer de datos de rayos X, se pueden generar a partir de átomos y/o fragmentos moleculares con la ayuda de un programa de construcción de moléculas adecuado, o bien modificando una estructura similar existente en las bases de datos, sin olvidar que la forma de la molécula que lleva a cabo sus funciones en el contexto biológico no tiene por qué ser la forma estable que adopta en el estado sólido o en disolución.

Por otro lado, las técnicas encaminadas a la determinación de estructuras tridimensionales se han sofisticado hasta el extremo de que hoy día puede resultar posible

seleccionar una diana macromolecular adecuada, producirla en cantidades relativamente grandes, purificarla, y estudiarla mediante técnicas de RMN, o bien cristalizarla para su análisis por difracción de rayos X. Cuando esto no es posible se puede recurrir a su modelado, teniendo en cuenta que la predicción de la estructura de un receptor proteico se enfrenta con dos problemas básicos:

1) el espacio configuracional asequible a una proteína, es decir, el número de conformaciones accesibles a una cadena polipeptídica, es enorme. Por ejemplo, una proteína de tan sólo 100 residuos de aminoácidos posee unos 300 grados de libertad de ángulos torsionales. Si cada ángulo torsional tiene acceso a, digamos, tres valores de mínima energía, la cadena polipeptídica puede dar lugar al menos a 3^{300} conformaciones. Este vasto espacio configuracional a explorar se puede reducir si se dispone de información procedente de métodos experimentales (*e.g.* RMN o datos cristalográficos de baja resolución) (van Gunsteren, 1988).

2) la descripción teórica de la interacción entre los átomos de una proteína requiere la formulación precisa de un potencial o campo de fuerzas empírico que contenga términos para representar los enlaces covalentes, los ángulos de enlace y torsionales, los enlaces de hidrógeno, y las interacciones electrostáticas y de van der Waals. Los campos de fuerzas actualmente en uso no han alcanzado todavía el nivel de sofisticación necesario para predecir con precisión la estructura de una proteína a partir únicamente de su secuencia primaria.

A pesar de lo anteriormente expuesto, los métodos computacionales de predicción de estructuras de proteínas, tanto estáticos como dinámicos, son muy útiles y su poder predictivo varía en función de la complejidad del sistema sometido a estudio. Podemos distinguir tres niveles de dificultad, en relación inversa con su exactitud relativa:

7.1. Modificaciones locales de estructura: a partir de la estructura conocida de una

proteína, se intenta predecir la estructura y propiedades de otra u otras proteínas que difieren tan sólo en unos pocos aminoácidos (Ortiz, 1992). Para ello se sustituyen las cadenas laterales de los aminoácidos diferentes, optimizando su orientación mediante una búsqueda sistemática de conformaciones de baja energía seguida de optimización del entorno del nuevo aminoácido mediante un algoritmo de minimización de la energía (Schiffer, 1990). La configuración de mínima energía resultante se encontrará generalmente próxima a la inicial puesto que el minimizador mueve al sistema básicamente "cuesta abajo" en la hipersuperficie energética del espacio configuracional, de modo que se ignoran posibles respuestas estructurales de la proteína a la sustitución del aminoácido. La búsqueda en el espacio configuracional se puede expandir haciendo uso de simulaciones de dinámica molecular (DM), que resuelve las ecuaciones clásicas de Newton del movimiento para todos los átomos del sistema y permite la superación de barreras de energía. La DM explora así una parte mucho mayor del espacio configuracional y permite localizar mínimos energéticos más bajos que los alcanzados mediante el uso de los minimizadores de energía ordinarios.

7.2. Modelado de proteínas homólogas: el plegado global del esqueleto polipeptídico es el mismo pero sólo se conserva un pequeño porcentaje de residuos de aminoácidos, pudiendo jugar un papel fundamental en el mantenimiento de la estructura los efectos del disolvente (Novotny, 1984). Este tipo de modelado puede ser muy ventajoso para guiar el diseño de nuevos ligandos. Así, el sitio activo de la carboxipeptidasa A bovina, cuya estructura tridimensional era conocida, sirvió para definir las analogías existentes entre esta proteína y la enzima convertidora de angiotensina (ECA), y para caracterizar la naturaleza y geometría de los grupos funcionales que deberían participar en la unión de sustratos peptídicos. Este modelo simplificado se empleó ventajosamente para diseñar potentes inhibidores que condujeron al desarrollo de antihipertensivos de la familia del

captoprilo (Cushman, 1977). La renina fue una de las primeras enzimas modeladas en su totalidad. Su modelado tridimensional se basó en la estructura conocida de la endothiapepsina y la secuencia primaria de la renina de ratón, muy semejante a la de la enzima humana (Blundell, 1983). Más recientemente, se han conseguido avances considerables en la creación de modelos tridimensionales fiables de receptores farmacológicos que funcionan acoplados a proteínas G (Trumpf-Kallmeyer, 1992) basados en la estructura de la bacteriorrodopsina, determinada por criomicroscopía electrónica.

7.3. Predicción de estructuras 'de novo': la proteína se considera como un conjunto de elementos de estructura secundaria (hélices α , láminas β , giros β , etc), cuyo ensamblaje da lugar a la estructura terciaria. Aunque se han hecho grandes esfuerzos para predecir la estructura secundaria a partir de la secuencia de aminoácidos, ninguno de los métodos actuales (*e.g.* Garnier, Chou-Fasman) ni siquiera identifica correctamente todos los elementos de estructura secundaria.

8. MÉTODOS DE SIMULACIÓN PARA SISTEMAS MOLECULARES.

Entre los métodos teóricos utilizados hoy día para el refinamiento de modelos moleculares y para el estudio de las interacciones fármaco-receptor se incluyen las minimizaciones de energía (ME), las simulaciones de dinámica molecular (DM), las simulaciones de Monte Carlo (MC), y las perturbaciones de la energía libre (PEL). Todas estas técnicas ofrecen en última instancia la posibilidad de comprender el comportamiento de péptidos, proteínas, y ácidos nucleicos, y el de sus complejos entre sí o con moléculas orgánicas, en términos de fuerzas moleculares fundamentales. Para ello dependen de la fiabilidad de la representación analítica de la hipersuperficie energética de las moléculas,

que viene dada por la suma de una serie de funciones simples que expresan tanto la energía requerida para distorsionar las coordenadas internas más allá de unos valores de equilibrio como las interacciones entre átomos no enlazados.

La función energética que describe la energía potencial del sistema en función de las posiciones de los átomos recibe el nombre de *campo de fuerzas*, y presenta una forma semejante a la ecuación [3], tomada del programa AMBER (Weiner, 1986), que se encuentra parametrizado para proteínas y ácidos nucleicos.

$$V(R) = V_b + V_\theta + V_\phi + V_{VDW} + V_{ELE} + V_{EH} =$$

$$\sum_{n=1}^{N_b} K_{b,n} (b_n - b_{eq,n})^2 + \sum_{n=1}^{N_\theta} K_{\theta,n} (\theta_n - \theta_{eq,n})^2 +$$

$$\sum_{n=1}^{N_\phi} \sum_{m=1}^{N_\eta} \frac{V_{m,n}}{2} [1 + \cos(\eta_{mn} \cdot \phi_{mn} - \gamma_{m,n})] +$$

$$\sum_{j=1}^{N_{tomos}} \sum_{i>j}^{N_{tomos}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) +$$

$$\sum_{j=1}^{N_{enlaces H}} \sum_{i>j}^{N_{enlaces H}} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) +$$

$$\sum_{j=1}^{N_{tomos}} \sum_{i>j}^{N_{tomos}} \frac{q_i \cdot q_j}{\epsilon \cdot r_{ij}} \quad [3]$$

$$\text{donde: } r_{ij}^* = r_i^* + r_j^*$$

La suma de términos incluye interacciones enlazadas (distancias, b , y ángulos, θ y

ϕ) y no enlazadas (van der Waals, electrostáticas, y enlaces de hidrógeno). La energía es función del conjunto de coordenadas cartesianas, \mathbf{R} , que especifican las posiciones de todos los átomos implicados, de una serie de parámetros energéticos (A, B, C, D, y las constantes de fuerza K), de las cargas atómicas (q), de la constante dieléctrica del medio (ϵ), y de los valores geométricos de equilibrio. La exactitud de los resultados de las simulaciones dependerá de la precisión del método de derivación de los parámetros utilizados y de la forma analítica de representar la energía. Las limitaciones y deficiencias de los campos de fuerza utilizados se ponen de manifiesto cuando se comparan los resultados de la simulación con datos experimentales.

La expresión de la energía de la ecuación [3] es tratable computacionalmente sólo para sistemas con un número de átomos relativamente pequeño. El número de coordenadas internas crece linealmente con el tamaño del sistema molecular, de modo que el tiempo de cálculo invertido en evaluar los tres primeros sumatorios también crece de forma linear. Sin embargo, la inspección de los tres sumatorios siguientes, que representan las interacciones no enlazadas, revela una dependencia cuadrática del número de átomos del sistema. Por esta razón es corriente introducir una distancia límite (*cut-off*) para que el algoritmo no considere las interacciones no enlazadas entre aquellos átomos separados entre sí por distancias mayores de un valor predeterminado. De este modo se acortan sustancialmente los tiempos de cálculo, aunque se corre el riesgo de que las discontinuidades creadas en la energía y sus derivadas conduzcan a artefactos en las minimizaciones y a la no conservación de la energía en las trayectorias de dinámica molecular. En la práctica, los programas crean una lista de aquellos pares de átomos cuya distancia entre sí es menor que la distancia límite especificada por el usuario, y la actualizan cada cierto número de pasos de minimización o de dinámica. Otra consideración importante a la hora de definir esta selección de átomos concierne a las interacciones electrostáticas. La utilización de distancias límite para el potencial de

interacción de van der Waals es bastante razonable, puesto que es un potencial de corto alcance y se extingue en función de $1/r^6$. A unos 8-10 Å de distancia, que son algunos de los límites corrientes, la energía y las fuerzas son bastante pequeñas, de modo que igualar a cero este potencial a una distancia de 10 Å es una aproximación razonable. Por el contrario, las interacciones de Coulomb se extinguen en función de $1/r$, lo que significa que incluso a distancias considerables este término no es despreciable. Sin embargo, si exceptuamos unos pocos grupos con carga formal, la mayoría de las moléculas están compuestas de fragmentos neutros con dipolos y cuadrupolos, por lo que en realidad el término de mayor importancia en las interacciones electrostáticas entre moléculas o entre partes de una misma molécula es la interacción dipolo-dipolo, que decrece en función de $1/r^3$. Un ejemplo sencillo puede servir para comprender las implicaciones de este razonamiento: la interacción entre dos monopolos, cada uno con una unidad electrónica de carga, es de unas 33 kcal mol⁻¹ a 10 Å, mientras que la energía de interacción entre dos dipolos formados a partir de monopolos unidad no es mayor de 0.3 kcal mol⁻¹. De aquí se desprende que ignorar la interacción monopolomonopolo conduciría a un error considerable, en tanto que ignorar la interacción dipolodipolo se puede considerar una aproximación modesta. Por todo lo anterior queda claro que la aplicación de un *cut-off* sobre la base de átomos individuales puede fácilmente conllevar la separación artificial de dipolos (dejando a un átomo dentro de la distancia límite y a otro fuera), lo que supone introducir una interacción monopolomonopolo considerable, en lugar de ignorar una interacción dipolodipolo relativamente pequeña. Por ello es preferible definir el *cut-off* sobre la base de residuos o grupos de átomos cuya carga neta sea cero o casi cero.

8.1. Minimización de la energía.

La minimización de la energía de pequeñas moléculas o de sistemas

macromoleculares se lleva a cabo por cálculos iterativos sucesivos en los que se somete a la conformación inicial del sistema a un proceso de optimización geométrica parcial o completa. Todos los parámetros que definen la geometría del sistema, o en su caso sólo algunos, se modifican en pequeños incrementos hasta que la estructura alcanza un mínimo energético local. Aunque algunos algoritmos hacen uso de la derivada de la función (gradientes conjugados), o incluso de la derivada segunda (Newton-Raphson), para guiar la minimización, ninguno de estos métodos puede garantizar que se ha encontrado la estructura de energía más baja posible, el llamado *mínimo global*. La minimización puede proceder bien en coordenadas internas (enlaces y ángulos) o, más frecuentemente para sistemas macromoleculares, en coordenadas Cartesianas (cada átomo está caracterizado por unas coordenadas x , y y z , y el átomo se mueve con pequeños incrementos a lo largo de estos ejes). Una de las ventajas de minimizar en coordenadas internas es que se pueden simular bien los movimientos cooperativos de varios átomos o grupos de átomos y se reduce el riesgo de que las moléculas queden atrapadas en falsos mínimos.

Las minimizaciones se pueden controlar mediante el uso de restricciones. Por ejemplo, se pueden fijar las coordenadas de algunos átomos, o constreñir la distancia entre dos átomos dados, enlazados o no, a un valor prefijado aplicando una constante de fuerza adecuada. Igualmente se pueden fijar restricciones a las torsiones alrededor de determinados enlaces o forzar a una molécula a que adopte la conformación de otra ("forzado sobre plantilla") (Baniak, 1987).

8.2. Análisis conformacional.

En un problema con múltiples variables, como es el de encontrar el conformero de mínima energía de una molécula o la conformación más estable de un complejo, la respuesta que se obtiene depende en gran medida del punto de partida. Además, la mayoría

de las veces estamos interesados no sólo en el mínimo global sino en todos aquellos conformeros que pueden estar poblados de forma significativa a temperatura ambiente, o hasta una pocas kcal/mol por encima del mínimo global. Para localizar todas estas conformaciones de baja energía definidas por una compleja hipersuperficie de energía potencial es necesario realizar una *búsqueda conformacional*. El punto de partida lo puede constituir un conformero cualquiera, que se optimiza mediante una técnica de minimización de la energía, y se compara con otros conformeros encontrados previamente para chequear su posible duplicación. Si el nuevo conformero así generado no había sido descubierto aún, se añade a una lista acumulativa de conformeros distinguibles entre sí. Este ciclo se continúa utilizando una nueva geometría de partida, minimizando, etc. Llegará un momento en que se habrán utilizado todas las geometrías iniciales posibles o en que no se encuentren nuevos mínimos, con lo que se dará por finalizada la búsqueda. Como la parte correspondiente a la minimización simplemente refina las estructuras de partida hasta alcanzar el mínimo local más próximo, quien más controla la eficacia de la búsqueda para alcanzar la convergencia es el algoritmo que genera los conformeros iniciales.

Los métodos utilizados para generar un conjunto de configuraciones del sistema molecular se pueden dividir en tres grandes categorías: (1) búsquedas deterministas que cubren todas las áreas del espacio conformacional de forma sistemática, (2) métodos estocásticos (Monte Carlo) que utilizan un elemento aleatorio para la exploración del espacio, y (3) dinámica molecular. Los dos primeros pueden operar bien en coordenadas externas (Cartesianas), en coordenadas internas (enlaces y ángulos de enlace y torsionales), o en representaciones de geometría de distancias (matrices de distancias internucleares). Como el número de estructuras tridimensionales posibles se expande geométricamente con el número de grados de libertad, y es bien sabido que los distintos conformeros difieren principalmente en sus ángulos torsionales, es muy ventajoso y eficaz confinar la

búsqueda exclusivamente al espacio torsional. De este modo un problema en principio R^{3N} queda reducido como mucho a otro $R^{N'}$, siendo R el número de puntos muestreados a lo largo de la coordenada torsional, N el número total de átomos y N' el número de torsiones independientes (Chang, 1989; Saunders 1990). Una búsqueda conformacional de este tipo puede considerarse como una búsqueda "en árbol" (*tree search*), donde los grados de libertad sucesivos son los nodos del árbol y el muestreo produce las ramas que emergen de estos nodos (Figura 3). El tiempo necesario para llevar a cabo una búsqueda es proporcional al número de nodos en el árbol. Este tipo de método se encuentra implementado en el programa CONGEN (Brucoleri, 1987) para muestrear uniformemente el espacio conformacional de pequeños segmentos polipeptídicos de proteínas y así predecir el plegado de las cadenas laterales de los aminoácidos cuando se lleva a cabo el modelado por homología.

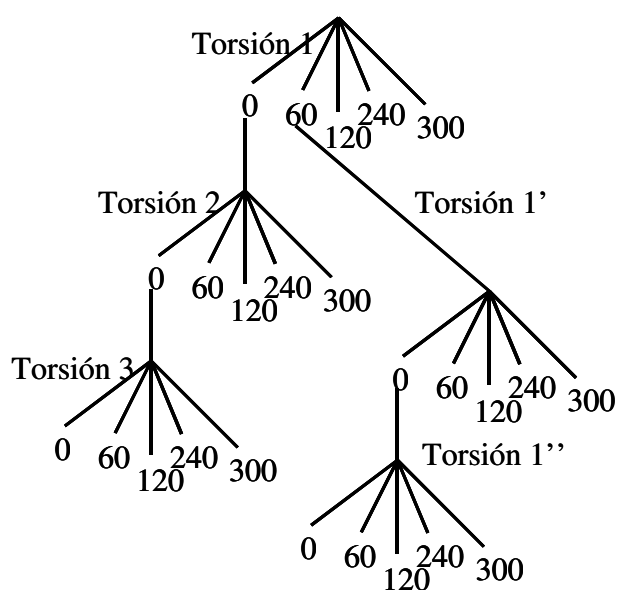


Figura 3. Porción superior de una árbol de búsqueda, que muestra cómo se puede representar el procedimiento de muestreo.

Una limitación de la mayor parte de las técnicas de muestreo actuales es que asumen que las cargas atómicas calculadas para una determinada geometría son válidas para todas las geometrías. En este sentido, es importante que las cargas utilizadas sean representativas de un amplio abanico de conformaciones (Reynolds, 1992).

8.2.1. Búsqueda sistemática ("*grid search*").

Si el sistema contiene sólo unos pocos grados de libertad se puede explorar la totalidad del espacio configuracional, variando cada uno de los torsionales de forma sistemática y minimizando la energía de cada uno de los conformeros así generados. Para una molécula con 3 ángulos dihedros, y un incremento del valor de cada ángulo de 60° , se generarán $6^3 = 216$ configuraciones. Si el número de torsionales considerados es 6, esta cifra aumenta a 46.656, y para 9 torsionales, el número de configuraciones generadas es superior a diez millones. Estos números sirven para poner de manifiesto que, al ser un procedimiento combinatorio, la demanda computacional restringe su uso a problemas con un número limitado de grados de libertad torsional y obliga en muchos casos a establecer un medio adicional de seleccionar sólo aquéllos conformeros que son únicos y de energía relativamente baja. La mayor parte de las mejoras a este método se centran en algoritmos de filtrado que permiten desconsiderar porciones poco razonables del espacio conformacional (Howard, 1988). Además hay que considerar que si el número de valores para cada torsión es pequeño, la resolución con que se cubre el espacio puede ser inadecuada para proporcionar una geometría de partida en la vecindad de cada mínimo, y se pueden perder mínimos de baja energía.

Una vez encontradas las distintas conformaciones posibles y ordenadas de acuerdo con sus energías, es de esperar que aquéllas con las energías más bajas sean las que

contribuyen a las propiedades del estado fundamental de la molécula. Si bien esto es cierto para algunas moléculas, no se puede decir lo mismo cuando las especies están cargadas o la molécula es polarizable. En este caso, es muy probable que las conformaciones más favorecidas sean aquéllas que maximicen las interacciones electrostáticas y el número de enlaces de hidrógeno. Tampoco hay que olvidar que si las energías de estos mismos conformeros se evaluaran en presencia de un disolvente polar, en lugar de *in vacuo*, un número significativo de esas interacciones serían reemplazadas por interacciones más favorables con moléculas del disolvente (Howard, 1988).

8.2.2. Monte Carlo (MC).

Si el sistema contiene muchos grados de libertad, se puede generar un conjunto de configuraciones utilizando una combinación de muestreo aleatorio y el empleo del factor de Boltzmann:

$$\exp[-V(r_1, r_2, \dots, r_N)/k_B T] \quad [4]$$

Dada una configuración de partida elegida arbitrariamente, se genera una nueva por desplazamiento al azar en espacio Cartesiano de uno o más átomos (o por variaciones en ángulos torsionales seleccionados), de forma tal que al cabo de un gran número de desplazamientos sucesivos se haya muestreado de forma uniforme el espacio disponible a todos los átomos. Cada nueva configuración se acepta o rechaza en base a un criterio energético (Metropolis, 1953) que implica un cambio en la energía potencial del sistema ($\Delta V(\mathbf{r})$) con respecto a la configuración anterior: se acepta si $\Delta V \leq 0$, o si para $\Delta V > 0$, $\exp(-\Delta V/k_B T) > R$, donde R es un número aleatorio tomado de una distribución uniforme sobre el intervalo (0,1). En caso de ser rechazada la actual, se vuelve a la configuración previa, que se utiliza de nuevo como punto de partida para otro desplazamiento al azar. De esta

manera cada configuración aparece con una probabilidad proporcional a su factor de Boltzmann y, si el paso básico se repite muchas veces, equivale a simular el movimiento de los átomos en contacto térmico con un baño de calor a la temperatura T . De este modo el sistema evoluciona a una distribución de Boltzmann. Para obtener una buena eficacia computacional sería de desear la combinación de un tamaño de paso (aleatorio) grande con una razón de aceptación alta. Por desgracia, para sistemas complejos que cuentan con muchos átomos enlazados covalentemente, sólo se puede obtener una razón de aceptación razonable para un tamaño de paso muy pequeño, lo cual limita la eficacia de este método para sistemas macromoleculares.

La muestra de Monte Carlo es una mera colección de estructuras tridimensionales arbitrarias, no relacionadas entre sí. Las configuraciones secuenciales de este tipo de muestreo son ' instantáneas' tomadas al azar (de ahí que el nombre del método haga referencia al célebre Casino) del movimiento total de la molécula y, por tanto, no ofrecen ninguna pista sobre la ruta traslacional que las une. No obstante, se puede extraer información respecto a la flexibilidad de la cadena a partir de las colecciones de datos conformacionales acumulados durante el proceso de muestreo, haciendo uso de la mecánica estadística. Las estructuras se pueden organizar sobre la base de algún criterio y compararse luego frente a modelos ideales o modelos de otras moléculas.

La eficacia de las técnicas de MC (definida en términos del cociente entre el número de estructuras aceptadas y el número total de tanteos) está determinada tanto por la temperatura, T , a la que tiene lugar el muestreo (cuanto más alta, mayor será la probabilidad de aceptación de la nueva estructura) como por el tipo y magnitud de los cambios conformacionales introducidos. Cambios demasiado grandes darán lugar a tasas de aceptación bajas por originar con facilidad solapamientos estéricos; ajustes demasiado pequeños harán necesarios muchos pasos "cuesta arriba" para cruzar una barrera de energía conformacional.

8.2.3. Dinámica molecular (DM).

A temperatura ambiente, un sistema molecular no es estático sino que atraviesa múltiples mínimos de la superficie de energía potencial. Cada átomo puede ser considerado como una partícula cuyo movimiento está determinado por las fuerzas que sobre ella ejercen todos los otros átomos de su entorno, según prescribe la física newtoniana ($F = m \cdot a$). Dadas las posiciones y velocidades de las partículas, así como la magnitud y dirección de las fuerzas que actúan sobre cada una de ellas en un instante determinado, por integración de las ecuaciones del movimiento de la mecánica clásica se pueden obtener la posición y la velocidad en un instante ligeramente posterior. Tomando pasos de tiempo sucesivos se puede construir una trayectoria dependiente del tiempo para todos los átomos, que representará así el movimiento de los mismos.

El movimiento de un átomo dado a lo largo de una coordenada determinada, por ejemplo x , se puede calcular de la siguiente forma: conocida su posición a tiempo t , $x(t)$, su posición tras un pequeño intervalo de tiempo, Δt , vendrá dada por la serie de Taylor mostrada en la ecuación [5].

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt} \Delta t + \frac{d^2 x(t)}{dt^2} \frac{\Delta^2 t}{2} + \dots \quad [5]$$

La solución numérica a las ecuaciones del movimiento consiste en conocer la posición $x(t)$, la velocidad $dx/dt(t)$ y la aceleración $d^2x/dt^2(t)$, y en hacer aproximaciones adecuadas para tener en cuenta las contribuciones de las derivadas más altas, de modo que $x(t + \Delta t)$ pueda calcularse con una precisión razonable. Para ello, el intervalo de integración ha de ser lo suficientemente breve (del orden de 0.5-1 femtosegundos) para que la fuerza que actúa en la nueva posición no difiera mucho de la que se registraba en la

posición original. El cálculo de esta nueva posición al final de este incremento de tiempo completa un ciclo del procedimiento que, repetido iterativamente, produce la trayectoria total en la forma de una serie de posiciones y velocidades en tiempos sucesivos. El conjunto de configuraciones así generadas se asume que constituyen una representación adecuada de la función de distribución completa.

El lado derecho de la ecuación [5] es una serie infinita que es necesario truncar para poder determinar $x(t + \Delta t)$. La forma de llevar esto a cabo constituye una de las diferencias fundamentales entre los diferentes algoritmos de DM, cuya eficacia se puede probar haciendo uso de la tercera ley de Newton, según la cual la fuerza neta que actúa sobre la totalidad del sistema de partículas es 0, de tal modo que deben conservarse la energía total del sistema (potencial + cinética) y el momento (McCammon, 1987b).

El paso de tiempo, Δt , es el principal factor limitante de la duración de la simulación. Como se encuentra restringido a tiempos menores que los movimientos de frecuencia más alta, *i.e.* el estiramiento y acortamiento de los enlaces, es frecuente utilizar algoritmos, *e.g.* SHAKE (Ryckaert, 1977), que constriñen las longitudes de enlace a sus valores de equilibrio, permitiendo así el uso de pasos de tiempo de hasta 2 fs.

Hoy día, los cálculos de DM incluyen moléculas de disolvente (generalmente agua) y contraiones, en un intento de modelar las moléculas en disolución de la forma más realista posible. La omisión del agua en el modelo puede conllevar el enterramiento de grupos cargados dentro de la estructura de una proteína y una distorsión muy marcada de estructuras polianiónicas, como son los ácidos nucleicos, fenómenos que no es probable que ocurran en disolución. Como los sistemas modelados contienen normalmente unos pocos miles de átomos, como máximo, mientras que los sistemas que pretenden representar poseen dimensiones prácticamente infinitas, es convencional crear copias del modelo (generalmente, un paralelepípedo rectangular), que se repiten periódicamente en todas las direcciones. De esta forma, cada paralelepípedo es simplemente una imagen del

sistema, que se encuentra así en las denominadas *condiciones de límite periódico*. Esto garantiza que todos los átomos se encuentren rodeados de átomos vecinos, bien 'reales' (los del modelo original) o imágenes de los mismos (los que se encuentran más allá de los límites), y que cuando una molécula abandona la caja por una cara, se introduzca de nuevo en ella por la cara opuesta, de modo que el número total de partículas se mantenga constante.

Desde el punto de vista de la mecánica estadística, una simulación de macromoléculas *in vacuo* o una simulación que incluya el disolvente y condiciones fijas de límite periódico corresponde a un 'conjunto o ensamblado microcanónico', en el que la energía total del sistema se conserva, así como el volumen y el número de partículas. Los sistemas se suelen simular en condiciones de temperatura y volumen constantes ('conjunto canónico' o de temperatura y presión constantes ('conjunto isotérmicoisobárico'), ya que éstas son las condiciones en las que se realizan la mayor parte de los experimentos.

Para comenzar una simulación de DM, es necesario disponer de un conjunto inicial de posiciones atómicas y de velocidades. La estructura de partida normalmente ha de ser refinada energéticamente, utilizando un algoritmo de minimización, para aliviar posibles tensiones locales derivadas de solapamientos atómicos o de distorsiones de enlaces o ángulos. Las velocidades iniciales se asignan a los átomos de forma aleatoria a partir de una distribución de Maxwell-Boltzmann correspondiente a una baja temperatura, y se realiza una primera simulación de DM durante unos pocos picosegundos. Le sigue un período de equilibrado durante el cual continúan alternándose nuevas asignaciones de velocidades, elegidas a partir de distribuciones Maxwellianas correspondientes a cada nueva temperatura sucesivamente aumentada, con intervalos similares de relajación dinámica hasta alcanzar la temperatura deseada. La temperatura, T , para este conjunto microcanónico se mide en términos de la energía cinética media para el sistema compuesto de N átomos como:

$$\frac{1}{2} \sum_{i=1}^N m_i \langle v_i^2 \rangle = \frac{3}{2} N k_B T \quad [6]$$

donde m_i y $\langle v_i^2 \rangle$ son la masa y el cuadrado de la velocidad media del átomo i , respectivamente, y k_B es la constante de Boltzmann. El período de equilibrado se considera terminado cuando no son evidentes cambios sistemáticos en la temperatura del sistema durante unos 10 ps.

La eficacia de la DM está determinada en gran medida por la temperatura, T , a la que tiene lugar el muestreo. Si T es demasiado alta, la molécula pasará demasiado tiempo en estados de alta energía y visitará los mínimos energéticos con muy poca frecuencia. A la inversa, si T es demasiado baja, la molécula será incapaz de superar barreras de energía con la rapidez necesaria para acceder a nuevas regiones del espacio configuracional en un tiempo razonable. La realización de simulaciones de DM a temperaturas altas conlleva un aumento del tamaño del espacio conformacional explorado, pero (sobre todo en el caso de macromoléculas) hay que tener en cuenta que la temperatura de la simulación está limitada por el requerimiento de mantener la estructura secundaria (evitar isomerizaciones *trans-cis* de enlaces peptídicos, desapareamientos de bases en el ADN, etc.), que puede hacer necesario en ciertos casos la imposición de restricciones. Generalmente las conformaciones generadas a altas temperaturas (500-1000 K) minimizan a energías por debajo de la inicial, demostrando la utilidad del método para superar barreras de energía alrededor del mínimo local (Brucoleri, 1990). Incluso a temperatura ambiente, la DM puede ayudar al refinamiento energético de una macromolécula, especialmente en el caso de estructuras procedentes de RMN o cristalografía de rayos X, donde ya están presentes restricciones adicionales (Brünger, 1991).

En el estudio de procesos dinámicos en los que existen barreras potenciales sustanciales, la DM puede resultar ineficaz para un muestreo adecuado con fines

estadísticos. Para muestrear grandes regiones del espacio conformacional mediante DM lo más adecuado es generar un cierto número de conformaciones iniciales diferentes, y utilizar cada una de ellas como punto de arranque para diferentes trayectorias (Howard, 1988).

El enfriamiento rápido de un sistema simulado a alta temperatura, *v.g.* por minimización de la energía, lo conduce al mínimo local más próximo, que puede ser un estado sobrecalentado. Para evitar el atrapamiento del sistema en "falsos mínimos" de alta energía, es útil enfriarlo y proseguir la dinámica a una temperatura más baja (*e.g.* 300 K) con objeto de que pueda escapar hacia mínimos de energía más bajos (Mackay, 1989; Ortiz, 1993). Otra alternativa es el llamado "*simulated annealing*", proceso durante el cual el sistema se calienta inicialmente a una alta temperatura para luego proceder a su enfriamiento lento hasta que se produce su "congelación" y ya no sufre más cambios (Kirkpatrick, 1983). A cada temperatura sucesivamente más baja la simulación debe proceder durante un tiempo suficientemente largo para que el sistema alcance un estado estacionario.

8.2.4. Híbrido DM/MC.

Uno de los principales inconvenientes de utilizar técnicas de DM para explorar el espacio conformacional es que gran parte de la energía cinética disponible se disipa en vibraciones de enlaces y ángulos, que son irrelevantes para la búsqueda conformacional. Las transiciones conformacionales que implican rotaciones concertadas alrededor de ángulos dihedros flexibles son relativamente infrecuentes. El algoritmo SHAKE (Ryckaert, 1977) intenta solventar este problema, eliminando estas vibraciones de las trayectorias de DM. También se pueden aplicar técnicas de filtrado de Fourier a los datos sin procesar de la dinámica para eliminar los movimientos indeseados y aislar las transiciones conformacionales de baja frecuencia (Dauber-Osguthorpe, 1990). Por su

parte, en las simulaciones de MC, al mantenerse completamente fijos las longitudes y los ángulos de enlace mientras se someten los dihedros a rotaciones aleatorias, las barreras rotacionales pueden llegar a ser artificialmente altas.

Recientemente se ha propuesto una solución intermedia, el Monte Carlo Dinámico (Morley, 1992), que aprovecha las ventajas respectivas de los dos métodos: se generan configuraciones de partida mediante cortas simulaciones de DM en las que la energía cinética se concentra en un enlace rotable seleccionado al azar. La trayectoria de DM continúa en tanto en cuanto las energías de las nuevas estructuras satisfagan la prueba de Metropolis, permitiendo así una rápida localización de regiones del espacio conformacional de baja energía. Si se rechaza la nueva configuración, se selecciona un nuevo enlace para empezar un nuevo ciclo de dinámica.

8.3. Método de la perturbación de la energía libre.

Los resultados del modelado y la minimización de la energía en el mejor de los casos son cualitativos debido a las limitaciones de los cálculos. De las trayectorias de DM se pueden obtener los valores promediados estadísticamente en el equilibrio para cualquier propiedad del sistema que sea medible directamente en cada punto de la trayectoria (*e.g.* la energía cinética de partes relevantes del sistema, propiedades estructurales, campos eléctricos). A partir de estos valores promedios se pueden derivar también un número importante de propiedades termodinámicas, aunque dos cantidades importantes, la entropía y la energía libre (de Helmholtz o de Gibbs) no se pueden derivar de un promedio estadístico, por ser propiedades globales que dependen de lo extenso que sea el espacio configuracional (o de fase) accesible al sistema molecular. Esto hace que sea virtualmente imposible calcular la energía libre absoluta de un sistema molecular. Sin embargo, la constante de unión de un ligando a su receptor está directamente relacionada con la energía libre (ecuación [1]), por lo que la determinación de esta cantidad es de sumo

interés. Afortunadamente, en los últimos años se han desarrollado una serie de procedimientos para evaluar cambios relativos de energía libre que abren perspectivas hasta hace poco impensables en el estudio de la interacción fármaco-receptor.

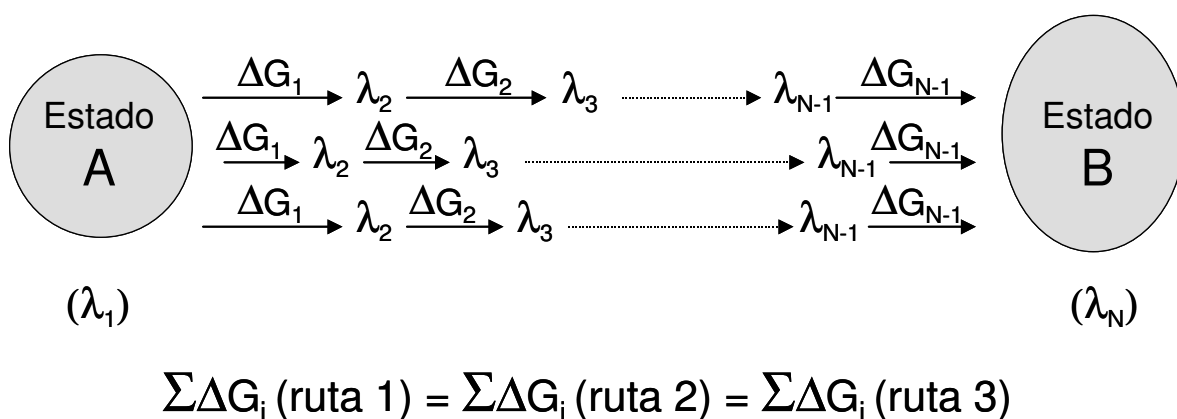


Figura 4. Rutas posibles de transformación del estado A en B a lo largo de una perturbación termodinámica.

Los métodos de perturbación e integración termodinámica se basan en que los *cambios* de energía libre relacionados con pequeñas perturbaciones de un sistema molecular se pueden determinar a partir de una simulación (de MC o DM). La diferencia de energía libre entre dos estados, A y B, de un sistema se puede determinar a partir de una simulación de DM en la que la función de energía potencial, $V(\mathbf{r})$, se altera lentamente de modo que el sistema A se va transformando poco a poco en el sistema B a lo largo de una ruta reversible (Figura 4). Para ello, el Hamiltoniano $H(\mathbf{p}, \mathbf{r})$, o sólo el término de energía potencial $V(\mathbf{r})$, se hace función de un *parámetro de acoplamiento*, λ , tal que $H(\mathbf{p}, \mathbf{r}, \lambda_A)$ caracteriza al estado A del sistema y $H(\mathbf{p}, \mathbf{r}, \lambda_B)$, al estado B. λ define una ruta de transformación entre los estados iniciales ($\lambda=0$) y finales ($\lambda=1$). Esta ruta no tiene por qué ser una ruta física si la cantidad de interés es una función de estado, como ΔG , que es

independiente del camino que siga la transformación. Esta es la ventaja crítica en la práctica pues la ruta no física puede ser computacionalmente ventajosa, de modo que se puede "mutar" un grupo funcional en el ligando o en el propio receptor. Algunos de los tipos de cambios que se pueden hacer en términos de "coordenadas de transición topográfica" aparecen esquematizados en la figura 5.

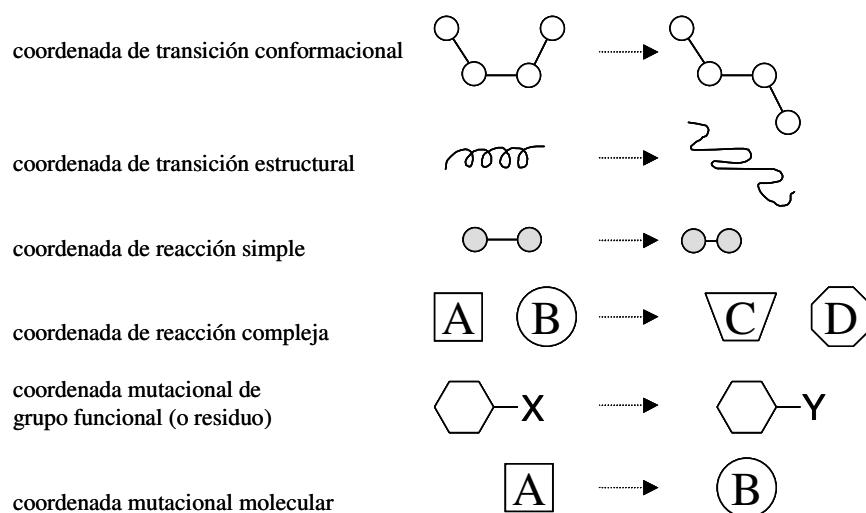


Figura 5. Tipos de coordenadas de transición topográfica utilizadas en las determinaciones de energías libres por medio de simulaciones moleculares.

8.3.1. Formulación del problema.

En un conjunto canónico (T , V y N constantes), la fórmula fundamental para la energía libre de Helmholtz es:

$$F = -k_B T \ln Z \quad [7]$$

donde k_B es la constante de Boltzmann, T es la temperatura absoluta y Z es la función de reparto, determinada por el Hamiltoniano $H(\mathbf{p}, \mathbf{q})$ que describe la energía total del sistema en términos de las coordenadas \mathbf{q} y los momentos \mathbf{p} de los átomos del sistema:

$$Z = [h^{3N} N!]^{-1} \int \exp[-H(p, q) / k_B T] dp dq \quad [8]$$

donde N es el número total de átomos y h la constante de Planck.

La correspondiente energía libre de Gibbs viene dada por:

$$G = F + PV = U + PV - TS \quad [9]$$

donde P es la presión del sistema y $U + PV$ es la entalpía.

La diferencia de energía libre, ΔG , entre dos estados bien definidos, A y B, cuyas funciones de reparto respectivas vienen dadas por Z_A y Z_B , se puede escribir como:

$$\Delta G = G_B - G_A = -k_B T \ln \frac{Z_B}{Z_A} \quad [10]$$

Aquí entra en juego el parámetro de acoplamiento, λ , para evitar tener que calcular independientemente Z_A y Z_B :

$$G(\lambda) = -k_B T \ln Z(\lambda) \quad [11]$$

Al ir variando λ desde 0 (donde Z es 100% Z_A) hasta 1 (donde Z será 100% Z_B) continua y lentamente, la energía potencial del sistema, $V(\lambda)$ va pasando lentamente de V_A a V_B . La diferencia de energía libre ΔG_{BA} viene dada por:

$$\Delta G_{BA} = G(\lambda_B) - G(\lambda_A) = -k_B T \ln \frac{Z(\lambda_B)}{Z(\lambda_A)} \quad [12]$$

que se puede expresar como un promedio de conjunto:

$$\Delta G_{BA} = -k_B T \ln \left\{ \exp \left[-\frac{H(p, r, \lambda_B) - H(p, r, \lambda_A)}{k_B T} \right] \right\}_{\lambda_A} \quad [13]$$

donde $\langle \dots \rangle_{\lambda}$ significa un promedio del conjunto sobre las coordenadas \mathbf{r} y los momentos \mathbf{p} al valor λ .

Para que esta *fórmula de la perturbación* dé resultados exactos, el estado B debe estar próximo al estado A. Normalmente la diferencia es lo suficientemente grande como para que sea necesario descomponer el cambio desde A hasta B en un número de pasos

entre estados intermedios que se encuentran lo suficientemente próximos para permitir el uso de esta fórmula (Figura 4). ΔG_{BA} viene dada entonces por la suma del conjunto de ΔG para cada paso.

Diferenciando en la ecuación [13] con respecto a λ a temperatura constante, T , se obtiene directamente:

$$\left[\frac{\delta G(\lambda)}{\delta \lambda} \right]_{T,P} = - \frac{k_B T}{Z(\lambda)} \left[\frac{\delta Z(\lambda)}{\delta \lambda} \right]_{T,P} = \left\langle \frac{\delta H(p, r, \lambda)}{\delta \lambda} \right\rangle_{\lambda} \quad [14]$$

que nos lleva a la *fórmula de la integración* para la diferencia de energía libre ΔG_{BA} :

$$\Delta G_{BA} = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\delta H(p, r, \lambda)}{\delta \lambda} \right\rangle_{\lambda} d\lambda \quad [15]$$

Si λ va cambiando muy lentamente desde λ_A hasta λ_B durante una simulación de dinámica molecular, la integración se puede ir haciendo a lo largo de la simulación. De este modo se puede obtener ΔG_{BA} para los estados algo diferentes A y B, con tal que el cambio continuo de λ sea tan pequeño que el sistema permanezca esencialmente en equilibrio para cada valor intermedio de λ .

En resumen, el cálculo de ΔG se puede ir haciendo bien integrando la derivada de $G(\lambda)$ a lo largo de λ (*integración termodinámica* o "crecimiento lento"), o bien designando estados intermedios $G(\lambda_i)$ que disten unos de otros suficientemente poco y calculando los cambios ΔG paso a paso (*perturbación termodinámica* o "ventanas"). Otra posibilidad sería ir desarrollando $G(\lambda)$ sobre el intervalo [A,B] a partir de simulaciones en las que λ es variable y obteniendo ΔG como $G(\lambda=1) - G(\lambda=0)$. La función $G(\lambda)$ es equivalente al "potencial de fuerza media" con respecto a la coordenada λ en termodinámica estadística de fluidos. Estos 3 tipos de cálculos constituyen hoy día las principales aproximaciones a

las determinaciones de la energía libre vía simulación molecular.

El factor crucial para obtener valores fiables de diferencias de energía libre (especialmente cuando contribuyen factores entrópicos) es el muestreo adecuado del espacio configuracional, que generalmente se lleva a cabo mediante simulaciones de DM. La energía del sistema se obtiene como un *promedio* sobre el conjunto de configuraciones generadas, mientras que la energía libre es una *integral* sobre el espacio configuracional, o de fase, al que es accesible el sistema. Una condición necesaria, aunque no suficiente, para obtener una estimación precisa de ΔG_{BA} consiste en realizar la integración en la ecuación [15] no sólo en la dirección desde el estado A al B, sino también en sentido contrario, desde B hasta A. Si el cambio de A a B a A se lleva a efecto de forma reversible, la histéresis será cero: $\Delta G_{BA} + \Delta G_{AB} = 0$. Otra condición a tener en cuenta es que el tiempo de simulación durante el cual se perturba el sistema desde A hasta B, y viceversa, debe ser mayor que el tiempo de relajación del sistema.

8.3.2. Ciclos termodinámicos.

Para utilizar las técnicas de perturbación o integración termodinámica para calcular energías libres relativas es necesario formular lo que se denomina un *ciclo termodinámico*. La clave del método radica en el hecho de que la energía libre es una función de estado, lo que significa que en tanto en cuanto el sistema se perturbe de forma reversible el cambio de energía libre, ΔG , será independiente de la ruta de perturbación elegida. Por lo tanto, a lo largo del ciclo o ruta cerrada, ΔG ha de ser igual a 0.

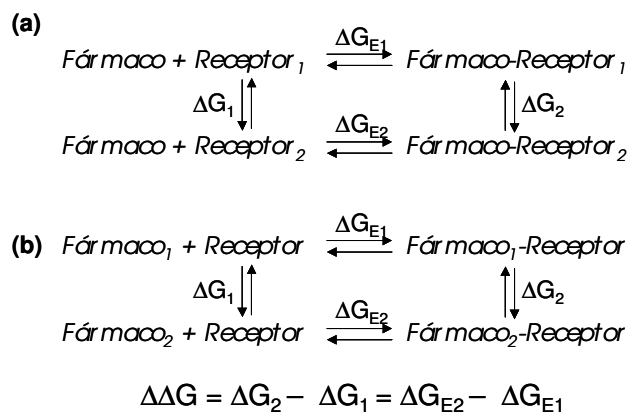


Figura 6. Ejemplos de perturbación de energía libre aplicados a la interacción fármaco-receptor.

Hay dos formas de obtener ΔG : una es calcular esta diferencia directamente a lo largo de la ruta que corresponde al proceso de interés; otra es diseñar un ciclo en el que el proceso que nos interesa sea sólo una parte, y calcular la diferencia de energía libre de la parte restante del ciclo. Aquí radica precisamente el poder del método, pues computacionalmente se pueden llevar a cabo procesos no físicos como, por ejemplo, la conversión de un átomo o un grupo químico en otro. Consideremos, por ejemplo, el proceso de unión de dos fármacos (G_A y G_B) a un receptor común R (Figura 6a). La simulación de los procesos que transcurren en sentido horizontal son prácticamente imposibles de simular pues implican la desolvatación del sitio receptor y de la molécula de fármaco para que pueda formarse el complejo fármaco-receptor. La constante de unión relativa vendría dada por:

$$\frac{K_B}{K_A} = e^{-[\Delta G_B - \Delta G_A]/RT} \quad [16]$$

donde R denota la constante de los gases. Sin embargo, la diferencia deseada no computable puede ser calculada indirectamente en virtud del ciclo termodinámico, ya que:

$$\Delta G_B - \Delta G_A = \Delta G_2 - \Delta G_1 \quad [17]$$

siempre que la composición del fármaco A no sea demasiado diferente de la del fármaco B. Esta técnica posibilita predecir de forma cuantitativa la manera en que un determinado sustituyente afectará a la afinidad de unión de un cabeza de serie al sitio receptor.

Esta metodología se ha aplicado a una variedad de pequeñas moléculas para calcular cambios relativos en su energía libre de solvatación (Charles, 1991) o en potenciales redox (Richards, 1989), y también a la interacción de inhibidores con enzimas, como por ejemplo la conversión de benzamidina a *p*-fluorobenzamidina en la tripsina (Wong, 1986) o la mutación NH → O en dos inhibidores peptídicos de la termolisina (Bash, 1987a). Igualmente resulta posible perturbar el sitio receptor (Figura 6b), por lo que también se han podido calcular los cambios en las propiedades de unión y catálisis de enzimas mutantes con respecto a las enzimas nativas (Bash, 1987b), y evaluar la diferencia en energía libre de unión de un antibiótico a dos secuencias diferentes de ADN (Gago, 1990).

8.3.3. Limitaciones del método PEL.

Aunque se han caracterizado algunos de los requisitos necesarios para juzgar si se ha alcanzado la convergencia, éstos pueden no ser suficientes para garantizar que se han muestreado todas las conformaciones accesibles que contribuyen al promedio del conjunto (Pearlman, 1989). El hecho de que el resultado final de los cálculos de PEL sea un simple número que tiene relevancia experimental ($\Delta\Delta G$) plantea la posibilidad de que el resultado obtenido sea el esperado pero de forma totalmente fortuita (o por compensación de errores, o por ajuste de parámetros). Son necesarias pruebas adicionales y valoraciones precisas de los efectos producidos por cambios en las condiciones de simulación (*cut-off*, formulación de r_{eq} para los átomos que 'desaparecen' o 'crecen', forma de cambio de ventanas modificadas dinámicamente, tratamiento de las condiciones de límite periódico, etc) para confirmar la validez de los resultados.

La fiabilidad de las predicciones va a depender también del sistema considerado y de la magnitud y naturaleza de los cambios introducidos. Si pensamos en un ligando que puede unirse a un receptor en dos orientaciones posibles con tan solo una pequeña diferencia de energía libre entre ambas, la sustitución de un grupo funcional por otro puede determinar una inversión en la preferencia orientacional del nuevo ligando con respecto al primero, y este cambio puede no ser observable en la simulación si la barrera de energía a salvar es considerable.

La energía libre de un sistema molecular es bastante sensible al potencial de interacción o campo de fuerzas utilizado en los cálculos. Para un caso relativamente sencillo como es la energía libre de solvatación del metanol en el agua, la aplicación de parámetros provenientes de distintos campos de fuerzas bien establecidos da lugar a una dispersión de unas $2kT$ alrededor del valor experimental de G (van Gunsteren, 1988). Esto quiere decir que los cálculos de la energía libre de solvatación de moléculas pequeñas, cuando están disponibles los valores experimentales, pueden proporcionar información sobre la validez de un campo de fuerzas y sobre cómo mejorar la calidad de los parámetros disponibles. Al mismo tiempo, esto quiere decir también que los resultados obtenidos de las simulaciones de PEL para interacciones ligando-receptor se deben chequear con respecto a variaciones en los parámetros del campo de fuerzas utilizado.

9.0 SELECTIVIDAD DE LA INTERACCIÓN FÁRMACO-RECEPTOR. "DOCKING".

A medida que se fueron cristalizando y analizando complejos macromolécula-ligando se fueron estableciendo las características generales de la unión. Pronto se puso de manifiesto que el ligando se mantiene en posición por muchas interacciones con diferentes grupos del receptor, grupos que en el caso de las proteínas pueden estar ampliamente dispersos en la secuencia de aminoácidos. Empezó a resultar aparente que una función

importante de la estructura macromolecular en su conjunto es precisamente mantener estos grupos en la orientación relativa adecuada para definir las características de la cavidad y constituir un sitio de unión específico. Las contribuciones individuales de cada grupo a la energía total de unión pueden ser débiles pero la suma de todas ellas puede hacer que el ligando se una muy firmemente. Como consecuencia de esta interacción pueden aparecer cambios alostéricos que, en el caso de las enzimas, o de la propia hemoglobina (Perutz, 1986), por ejemplo, son de fundamental importancia para comprender el mecanismo de acción de la proteína.

Uno de los ejemplos que mejor sirve para ilustrar el fenómeno de la selectividad es el de la dihidrofolato reductasa (DHFR), enzima diana para distintos agentes quimioterápicos, que cataliza la reducción de dihidrofolato a tetrahidrofolato, utilizando NADPH como cofactor. De los distintos inhibidores prototípicos, la pirimetamina inhibe selectivamente la enzima presente en *Plasmodium* sp., encontrando utilidad como antipalúdico, mientras la trimetoprima es más efectiva en bacterias. El metotrexato, por su parte, que muestra una constante de disociación parecida para todas las formas de la enzima, se utiliza para inhibir el crecimiento de células neoplásicas. Como se han conseguido analizar a alta resolución las estructuras cristalinas de varias DHFRs y se han preparado formas alteradas de la enzima mediante técnicas de mutagénesis dirigida, este sistema ha resultado singularmente propicio para su análisis desde un punto de vista tanto estructural como energético (Roberts, 1986). Igualmente, ha permitido examinar la especificidad de la interacción tanto en términos de la sustitución de determinados aminoácidos en la proteína como de la explotación de sitios de unión adicionales para mejorar la afinidad del inhibidor (Kuyper, 1985). Otro aspecto interesante que surgió como consecuencia de estos estudios es que el modo de unión de estos inhibidores está "invertido" con respecto al del propio dihidrofolato (Bolin, 1982), lo que acentúa la necesidad de prestar suficiente atención a las propiedades electrónicas de los ligandos y de

comprobar en qué medida los métodos teóricos son capaces de realizar este tipo de predicciones (Manaut, 1991).

En ausencia de datos cristalográficos sobre el modo preciso de unión del ligando a la macromolécula, es preciso explorar las diversas posibilidades de interacción entre ambas moléculas. Para que se forme un complejo, la energía de interacción (ΔE), ha de ser negativa:

$$\Delta E = E_{(\text{ligando-receptor})} - E_{\text{ligando}} - E_{\text{receptor}} \quad [18]$$

La computación de ΔE implica conocer tanto las conformaciones óptimas de ligando y receptor por separado como la conformación más estable del complejo. La búsqueda del complejo óptimo no es trivial en la mayoría de las ocasiones, y su modelado implica tener en consideración un buen número de factores, entre los que se encuentran:

a) conocimiento del sitio de unión óptimo en la macromolécula. En el caso de enzimas, el sitio receptor candidato va a ser el propio sitio activo, aunque en ocasiones el sitio relevante para la acción del fármaco puede ser diferente, lo que supone exploraciones adicionales (Ortiz, 1993).

b) determinación de la orientación relativa del ligando con respecto al sitio receptor, pues los casos de múltiples modos de unión están suficientemente documentados tanto para proteínas (Badger, 1988) como para ADN (Quintana, 1991). Los métodos teóricos normalmente buscan aquella orientación que da lugar a la energía de interacción más favorable, aunque cuando se ha podido comparar con resultados cristalográficos (Meng, 1992; Shoichet, 1993), se ha llegado a la conclusión de que los sitios calculados como más favorables no tienen por qué coincidir necesariamente con los observados experimentalmente.

c) caracterización de la conformación óptima del ligando en el sitio de unión.

d) evaluación de los posibles cambios conformacionales en el sitio receptor como consecuencia de la unión.

Las técnicas manuales de modelado del complejo ligando-receptor tratan inicialmente al sitio diana como si fuera completamente rígido, mientras que la conformación del ligando se va ajustando de forma interactiva, tanto por movimientos de traslación y rotación como mediante giros de enlaces rotables. Esta maniobra de acoplamiento ("*docking*", por analogía con la atracada de los buques en puerto) se simplifica considerablemente si se cuenta con información visualmente comprensible sobre las características químicas de ambas moléculas (*e.g.* superficies coloreadas codificadas según el potencial electrostático o hidrofóbico, o mapas energéticos de interacción con sondas de grupos funcionales). Los cálculos energéticos proporcionan asimismo una medida de la energía de interacción, que con algunos programas se puede monitorizar también interactivamente y sirve como guía para elegir la orientación preferida de una molécula respecto de otra. Pero aún utilizando una función simple de energía no enlazada que comprenda un componente estérico y otro electrostático, e ignorando las contribuciones entre átomos que pertenecen a una misma molécula (ecuación [19]), el tiempo de cálculo, proporcional al cuadrado del número de átomos, puede hacer que este proceso sea demasiado lento para sistemas moleculares grandes en la mayor parte de las estaciones de trabajo.

$$E_{interacción} = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon r_{ij}} \right) \quad [19]$$

La ecuación [19] contiene los términos *intermoleculares* presentes en la función de mecánica molecular de AMBER (ecuación [3]) con la excepción del término corrector del enlace de hidrógeno. Entre las simplificaciones propuestas está la de precalcular los términos de la función de energía potencial correspondientes a los átomos del receptor y

tabularlos en tres dimensiones (Pattabiraman, 1985). La aproximación utilizada para separar los términos de van der Waals consiste en calcular la media geométrica de A_{ij} y B_{ij} (Hagler, 1974):

$$A_{ij} = \sqrt{A_i A_j}; \quad B_{ij} = \sqrt{B_i B_j}$$

por lo que la ecuación [19] se puede reescribir como:

$$E_{interacción} = \sum_{i=1}^{lig} \sqrt{A_i} \sum_{j=1}^{rec} \frac{\sqrt{A_j}}{r_{ij}^{12}} - \sum_{i=1}^{lig} \sqrt{B_i} \sum_{j=1}^{rec} \frac{\sqrt{B_j}}{r_{ij}^6} + \sum_{i=1}^{lig} q_i \sum_{j=1}^{rec} \frac{q_j}{\epsilon r_{ij}} \quad [20]$$

Computacionalmente, el volumen ocupado por los átomos del receptor se encierra dentro de una malla tridimensional y para cada punto de la misma, m , se almacenan en memoria los tres sumatorios correspondientes a los átomos del receptor que se encuentren a menos de una distancia límite de ese punto. Para las interacciones electrostáticas, el uso de este *cut-off* debe tener en cuenta la no escisión del dipolo del residuo (*cf.* sección 8). Asumiendo que un punto de la malla se encuentra razonablemente próximo a un átomo de la molécula que se está sometiendo a movimiento, la cantidad r_{ij} se puede reescribir como r_{im} , que corresponde a la distancia entre el punto de la malla y el átomo de la molécula que se mantiene fija. Como quiera que r_{im} es constante (la malla no se mueve en relación con la molécula fija), los valores comprendidos en el interior de los sumatorios para la molécula i se pueden calcular una sola vez y almacenar en cada punto de la malla. La energía de interacción vendrá ahora dada por:

$$E_{interacción} = \sum_{i=1}^{lig} \left[\sqrt{A_i} M_A(\mathbf{P}_i) - \sqrt{B_i} M_B(\mathbf{P}_i) + q_i M_e(\mathbf{P}_i) \right] \quad [21]$$

donde $M_e(\mathbf{P}_i)$, $M_A(\mathbf{P}_i)$, y $M_B(\mathbf{P}_i)$ representan los valores en los puntos de la malla para las energías precalculadas electrostática (e), y Lennard-Jones repulsiva (A) y atractiva (B), siendo \mathbf{P}_j la posición del átomo en la molécula que se desplaza. Esta posición se utiliza

para asignar el valor de energía correspondiente al punto de la malla más próximo, directamente o por interpolación trilinear (Meng, 1992). Cuanto menor sea el ancho de la malla, más se aproximarán los resultados de la ecuación [21] a los resultados de la ecuación [20] ya que disminuye el error entre la posición atómica verdadera y el punto de la malla. Esto equivale a asumir que los átomos del ligando se encuentran siempre en puntos de esta red tridimensional, por lo que el error asumido en esta aproximación es la mitad del ancho de la malla. Un espaciado de 0.5 Å es adecuado para el cálculo del potencial electrostático, que es función del inverso de la distancia entre los átomos interactuantes, pero para el potencial de van der Waals, al ser función del inverso de la sexta y duodécima potencia de la distancia, es preferible utilizar un ancho de malla de 0.25 Å (Pattabiraman, 1985; Meng, 1992).

Como el proceso de *docking* no deja de ser un "rompecabezas tridimensional" (Kuntz, 1992), cuya resolución puede además verse sesgada por el propio usuario, se han desarrollado algunos métodos que intentan explorar las diversas alternativas de forma automatizada, muestreando los 6 grados de libertad implicados en la disposición espacial relativa de dos objetos tridimensionales rígidos. Así, el programa DOCK (Kuntz, 1982) primero caracteriza los surcos y las invaginaciones de la superficie de la macromolécula que pueden constituir sitios diana para el ligando (o sitios alostéricos o sitios sin función conocida), los cuales llena con conjuntos de esferas cuyos centros sirven para crear una imagen negativa del sitio. A continuación se utilizan algoritmos para ajustar las estructuras de los ligandos a la imagen del receptor sobre la base de una comparación de distancias internas y para realizar una superposición de mínimos cuadrados, evaluando todas las posibles orientaciones. La bondad del ajuste se mide mediante un método de puntuación simple de la complementariedad de forma (Kuntz, 1982) o, más recientemente, haciendo uso de un sistema de tanteo triple basado en contactos, potenciales electrostáticos y un campo de fuerzas intermolecular (Meng, 1992). Con este último marcador, que utiliza

también una malla tridimensional con valores precomputados para acelerar los cálculos, mejora mucho la calidad de las predicciones.

En otro procedimiento, desarrollado para acoplar sustratos en proteínas (Goodsell, 1990), se combina un método de exploración conformacional para el ligando con el cálculo de potenciales de afinidad en puntos de una malla tridimensional en la que se encuentra incluida la macromolécula. La malla de afinidad así computada para distintos tipos de átomos, así como una malla de potencial electrostático calculada mediante una carga puntual de +1, se almacena en memoria y se utiliza a modo de tabla para la evaluación rápida de la energía de interacción de cada conformero con la proteína.

Una vez modelados una serie de complejos, es necesario minimizar su energía, permitiendo la relajación de todos los átomos, para su comparación y para obtener una estimación de la entalpías relativas de interacción del ligando en cada uno de ellos, lo que proporciona una indicación de su verosimilitud. En este tipo de cálculos (realizados en el vacío) las interacciones iónicas y de enlace de hidrógeno están sobrestimadas porque las moléculas no compiten con el agua, que en la realidad actúa como disolvente y como aceptor/donador de enlaces de hidrógeno. No hay que olvidar que las energías de unión determinadas experimentalmente representan diferencias entre las energías de ligando y receptor unidos separadamente al agua y la de ligando y receptor unidos entre sí (Ward, 1990). Este proceso conlleva una pérdida de entropía que debe compensarse por la energía de unión que resulta de las interacciones no-covalentes favorables entre ligando y receptor (Page, 1977; Breslauer, 1987).

Los procedimientos de *docking* automático son preferibles a los interactivos para la búsqueda de prototipos en bases de datos, y pueden ser de utilidad para el diseño de nuevos ligandos (Shoichet, 1993).

10.0 DISEÑO DE NUEVOS LIGANDOS BASADO EN LA ESTRUCTURA TRIDIMENSIONAL DEL RECEPTOR.

Si la estructura de la macromolécula diana es conocida, se cuenta con una inestimable ayuda para generar ideas sobre posibles nuevos ligandos. Uno de los primeros ejemplos data de 1976, cuando las estructuras tridimensionales de la hemoglobina y sus complejos con el efector alostérico 2,3-difosfoglicerato inspiraron a algunos investigadores, utilizando modelos de varillas metálicas del sitio de unión, a diseñar ligandos con una estructura química completamente distinta y un modo de unión diferente que podían ser útil en el tratamiento de la anemia falciforme (Beddell, 1976).

Los modelos CPK o de varillas son adecuados hasta un cierto nivel de complejidad estructural, más allá del cual surgen problemas que hacen inabordable la exploración y manipulación sistemática sin que el modelo sufra una considerable degradación. Con los gráficos moleculares generados por ordenador, por otra parte, se pueden almacenar múltiples configuraciones del sistema, directamente relacionadas con listas de coordenadas, y no existe peligro de que el modelo se desarme. Adicionalmente, los programas de gráficos permiten una variedad de formas de representación, que no se limita a la forma molecular sino que incluye propiedades químicas de interés (Olson, 1992). La mera inspección visual de la superficie que envuelve a los átomos del sitio activo de una proteína puede ayudar a comprender el acoplamiento de ligandos e inhibidores. En algunos casos ha servido incluso para interpretar los coeficientes de los parámetros QSAR utilizados en las ecuaciones de correlación clásicas (Hansch, 1986). Si el sitio se examina con un ligando ya unido, se puede pensar en introducir variaciones modestas en el ligando conocido de modo que el ligando resultante sea de forma similar, pero con una afinidad aumentada. Alternativamente, las áreas y cavidades de la superficie que no estén ocupadas

serán indicativas de oportunidades para extender el ligando. En el caso de las enzimas, se puede pensar en estructuras que mimeticen el estado de transición predicho para un determinado sustrato, que en principio deberían unirse a la enzima más fuertemente que el propio sustrato, por lo que pueden resultar buenos inhibidores (Wolfenden, 1991).

Si se utiliza la superficie de Connolly o una superficie de van der Waals, los átomos del ligando se encuentran alejados un radio de van der Waals de la superficie de la proteína, lo que dificulta su examen. La superficie de Lee y Richards (1971), por el contrario, indica directamente las posiciones de los átomos de ligando que se encuentran en contacto de van der Waals. Si se representa el ligando en forma de esqueleto, todos los átomos que den lugar a buenos contactos de van der Waals se encontrarán muy próximos a la superficie, por lo que esta representación puede actuar de guía para mostrar dónde se deberían situar los átomos del ligando para que den lugar a contactos favorables. Si la superficie además proporciona información sobre la especificidad química de esa zona, puede encontrar aplicaciones interesantes como herramienta para el diseño de fármacos (Bohacek, 1992).

Para cuando no se cuenta con ningún cabeza de serie, o cuando se quieren diseñar compuestos con una estructura completamente diferente a la de los ligandos conocidos, se han desarrollado un cierto número de métodos que permiten proponer estructuras de moléculas con afinidad por un determinado sitio receptor de modo automático o semiautomático. Una primera posibilidad consiste en realizar una búsqueda sistemática en una base de datos de estructuras tridimensionales (*e.g.* la Base de Datos Estructurales de Cambridge, CSD (Allen, 1979) para identificar aquellos compuestos que pueden encajar en el sitio de interés. El programa DOCK (DesJarlais, 1988) fue una de las primeras contribuciones a este respecto. La principal ventaja de este método es que las moléculas investigadas existen realmente y se encuentran en una conformación de baja energía. El problema principal, sin embargo, es que no tiene en cuenta la flexibilidad conformacional,

y es un hecho bien conocido que la conformación del ligando unido no tiene por qué ser la misma que la encontrada en la estructura cristalina de la molécula aislada, por lo que cabezas de serie potenciales pueden pasar desapercibidos por encontrarse en una conformación inadecuada para la unión. Un método de estas características se encuentra además limitado por la naturaleza y número de las estructuras presentes en la propia base de datos. Más aún, este procedimiento considera al proceso de acoplamiento como una mera cuestión de geometría, dejando a un lado las propiedades químicas de los átomos individuales, las cuales van a ser determinantes a la hora de formar el complejo. Si las moléculas son rígidas, las geometrías de unión sugeridas por DOCK pueden corresponderse bastante bien con las halladas por cristalografía de rayos X, pero cuando hay cambios conformacionales implicados, la predicción se hace bastante problemática (Shoichet, 1993).

Para tener en cuenta tanto la forma como las características energéticas del sistema, conviene recordar que el potencial de Lennard-Jones viene expresado por la ecuación:

$$E_{LJ} = \frac{A}{r^{12}} - \frac{B}{r^6} \quad [22]$$

donde r es la distancia entre un par de átomos no enlazados cuya energía de Lennard-Jones, E_{LJ} , está descrita por los parámetros A y B . Para valores de E_{LJ} ligeramente negativos se puede encontrar un intervalo crítico de distancias interatómicas, por debajo del cual enseguida domina el término de repulsión, A/r^{12} , y por encima del cual la interacción tiende a desaparecer. Entre estos dos límites existe una zona de atracción mutua que puede representarse gráficamente como una superficie de contornos. Siguiendo este razonamiento, se podrían computar las interacciones de un átomo, o grupo funcional, con los átomos de una macromolécula diana en distintas posiciones, tanto a su través

como a su alrededor, y obtener así una serie de valores de energía para cada posición. Seleccionando unos niveles energéticos adecuados, se podrían representar gráficamente los contornos tridimensionales de las zonas donde ese átomo o grupo funcional daría lugar a una interacción favorable con la proteína. El programa GRID (Goodford, 1985) realiza justamente esta tarea, para lo que incluye además efectos electrostáticos y de enlace de hidrógeno. Aplicado a un receptor de estructura tridimensional conocida, permite predecir las interacciones no-covalentes que presenta con una serie de grupos funcionales o "sondas", caracterizadas por propiedades tales como radio de van der Waals, polarizabilidad, carga, hibridación, fuerza y longitud del enlace de hidrógeno, y número de enlaces de hidrógeno que es capaz de donar y aceptar. Esta descripción detallada es la que confiere propiedades específicas de interacción a cada sonda en GRID.

Es sabido que la interacción electrostática no disminuye rápidamente con la distancia y depende de forma crítica de la constante dieléctrica del medio. En GRID se asume que una interfaz plana separa una fase proteica homogénea de constante dieléctrica $\zeta=4$ de un disolvente homogéneo de constante dieléctrica $\epsilon=80$. La interacción electrostática disminuye progresivamente si la sonda o los átomos de la proteína se acercan a la superficie, y prácticamente desaparece a medida que la sonda entra en la disolución (la constante dieléctrica aumenta hasta $(\zeta + \epsilon)/2$). Los enlaces de hidrógeno son muy importantes para determinar la conformación de proteínas, polisacáridos, ácidos nucleicos, moléculas orgánicas y sus complejos respectivos. Se pueden considerar interacciones de alcance intermedio entre un átomo de hidrógeno deficiente en electrones y una región próxima de alta densidad electrónica. Tanto el átomo dador, que está covalentemente unido al hidrógeno, como el aceptor son altamente electronegativos, y cuando se forma el enlace de hidrógeno, la distancia entre ambos es menor que la suma de sus radios de van der Waals. Cuando la separación entre los átomos implicados es pequeña, el enlace de hidrógeno está gobernado por un equilibrio entre componentes

atractivos electrostáticos (transferencia de carga, polarización y dispersión), y términos repulsivos de intercambio de electrones. Cuando la distancia entre estos átomos es grande, el enlace de hidrógeno es casi por completo electrostático. Algunos potenciales de mecánica molecular incluyen un término adicional de enlace de hidrógeno, además de los términos electrostáticos y de van der Waals, para el ajuste fino de las distancias entre los átomos implicados (*e.g.* AMBER, ecuación [3]). En GRID también existe este potencial para el enlace de hidrógeno, y se formula de modo que la energía sea el producto de tres términos: E_r , una función 6-8 que depende de la separación entre los átomos, y E_t y E_p , que describen la dependencia orientacional de los enlaces de hidrógeno y dependen de la naturaleza química de los átomos y del ángulo que forma el enlace de hidrógeno con el átomo diana o con la sonda (Boobbyer, 1989). De los muchos enlaces de hidrógeno posibles que una sonda es capaz de establecer en una determinada posición *xyz*, se seleccionan los más favorables sin que se supere lógicamente el número máximo que los átomos de la proteína o de la sonda pueden donar o aceptar. Los átomos en GRID son representados como átomos "extendidos", es decir, esferas, centradas en una carga puntual, que consisten en un átomo más pesado que el hidrógeno y un número adecuado de hidrógenos que se consideran de forma implícita por modificación de los parámetros electrostáticos y de Lennard-Jones del átomo pesado. De este modo se tienen en cuenta tanto a los átomos de hidrógeno como a los pares de electrones sin compartir. Las funciones de enlace de hidrógeno para GRID han sido recientemente ampliadas a sondas capaces de formar dos (Wade, 1993a) o más (Wade, 1993b) enlaces de hidrógeno.

Una de las objeciones que se puede hacer a esta técnica es que en la computación de los contornos energéticos no se tiene en cuenta la entropía, que en la práctica es un factor muy importante a la hora de determinar la unión del ligando, puesto que la liberación de moléculas de disolvente ejerce una importante influencia en el efecto hidrofóbico (Tandford, 1980). Igualmente se puede argüir sobre la conveniencia de dividir

un ligando en sondas independientes o de asumir la aditividad entre pares de átomos. En cualquier caso, esta metodología permite estudiar detalladamente los atributos de un perfil energético en relación con la estructura de una proteína o un ácido nucleico, lo que puede resultar especialmente beneficioso cuando se intentan diseñar nuevos ligandos para que encajen en una determinada zona del receptor. El programa GRID permite representar gráficamente los sitios más favorables para cada tipo de sonda (Figura 7), junto con la penalización energética que se debe pagar si el grupo funcional en cuestión no se encuentra exactamente en el lugar óptimo. La proximidad de los contornos proporciona además información sobre la naturaleza y dirección de las fuerzas que actuarían sobre la sonda. La posibilidad de recontornear a diferentes niveles de energía permite filtrar los picos pequeños, conservando las características principales de la interacción predicha.

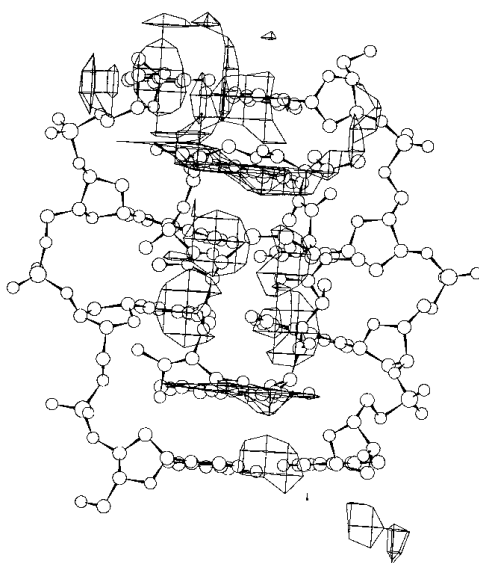


Figura 7. Mapas de GRID para una sonda de carbono sp^2 sobre un complejo ADN-fármaco bis-intercalante.

Los valores de interacción computados para cada sonda suelen diferir marcadamente. Es corriente encontrar sitios donde un grupo amino o carboxilo presenta

energías de interacción del orden de -15 a -20 kcal/mol, mientras que una sonda de metilo raramente consigue más de -4 kcal/mol, como consecuencia de su baja polarizabilidad y de no poder formar enlaces de hidrógeno. Esta información puede ser muy provechosa a la hora de diseñar ligandos de alta afinidad como posibles fármacos que contengan grupos amino o carboxilato, haciendo uso de los sitios de unión favorable presentes en proteínas o ácidos nucleicos que sirven como posibles dianas farmacológicas.

Otra característica notable de GRID es que puede utilizarse para la construcción de *mapas de selectividad*, que permiten comparar la interacción de una sonda con dos ligandos diferentes, por ejemplo, dos isómeros de una misma molécula. Para ello calcula la diferencia en energías de interacción de la sonda con cada isómero en todos los puntos de la malla tridimensional. Estas diferencias de energía se pueden asimismo contorneo a un nivel de energía adecuado para mostrar las zonas donde una determinada interacción está más favorecida en una molécula que en otra (Boobbyer, 1989). También es posible representar las diferencias en las energías de interacción entre dos sondas. Utilizando, por ejemplo, una sonda reducida y una sonda oxidada, el mapa resultante sería indicativo de aquellas regiones de la macromolécula que unirían un fármaco en su forma reducida, pero no en su forma oxidada, mostrando así sitios potenciales de unión para agentes biorreductores, teóricamente más selectivos frente a células tumorales pobremente oxigenadas (Reynolds, 1987).

En ocasiones puede no resultar fácil traducir a estructuras químicas concretas las energías de unión y las posiciones espaciales calculadas con GRID para una posible interacción. Cabría la posibilidad de ir posicionando fragmentos moleculares en determinadas posiciones del receptor, optimizando la formación de enlaces de hidrógeno y el relleno de cavidades hidrofóbicas, y conectarlos adecuadamente mediante otro tipo de

fragmentos "puente". Este procedimiento, al ser susceptible de automatización, puede resultar muy rápido, y presenta la ventaja de que puede generar una enorme variedad de moléculas, ya que el número de posibles combinaciones de fragmentos es muy grande. El programa LUDI (Böhm, 1992a,b) se desarrolló para llevar a cabo esta búsqueda semi-automatizada, en la que el investigador tiene capacidad de decisión. Para entender el funcionamiento de este programa resulta útil definir los términos *sitio de interacción*, *fragmento molecular*, y *puente*.

- Un *sitio de interacción* es una posición del espacio que no está ocupada por el receptor donde un átomo perteneciente a un grupo funcional del ligando puede establecer interacciones favorables con el receptor. LUDI distingue cuatro tipos diferentes de sitios de interacción: lipofílico-alifático, lipofílico-aromático, donador de enlace de hidrógeno y aceptor de enlace de hidrógeno. Los dos primeros son adecuados para dar lugar a interacciones hidrofóbicas. Los dos últimos, por el carácter fuertemente direccional del enlace de hidrógeno, se almacenan como vectores de modo que contengan información sobre su orientación.

- Un *fragmento molecular* es una molécula pequeña o un grupo funcional con geometría ideal que puede interaccionar favorablemente con alguno de los sitios de interacción potenciales.

- Un *puente* es también un grupo de átomos con geometría ideal pero que sirve para unir distintos fragmentos moleculares en una única molécula, o añadir un nuevo fragmento a un ligando ya existente.

Para correr el programa es necesario disponer de:

- a) las coordenadas cartesianas del receptor (incluyendo todos los hidrógenos de forma explícita) en formato PDB,
- b) un archivo de distribuciones de contactos no enlazados,
- c) una fragmentoteca (colección de fragmentos moleculares en formato PDB),

- d) una colección de "puentes" intramoleculares en formato PDB, y
- e) la posición del sitio receptor de interés (en forma de coordenadas para un determinado punto en ese sitio y una distancia límite como radio de acción).

En un primer paso, el programa calcula los sitios de interacción potenciales, para lo cual clasifica a los átomos del receptor en cuatro categorías: (i) donadores de enlace de hidrógeno (aquellos hidrógenos unidos covalentemente a átomos de N ó O), (ii) aceptores de enlace de hidrógeno (todos los átomos de nitrógeno que no están unidos covalentemente a hidrógenos, más todos los átomos de oxígeno), (iii) aromáticos (átomos de carbono de los anillos aromáticos de fenilalanina, tirosina, histidina y triptófano), y (iv) alifáticos: todos los átomos de carbono alifático.

Para generar los sitios de interacción, LUDI ofrece una de tres alternativas:

- 1) seguir un conjunto de reglas geométricas relativamente sencillas (Böhm, 1992b).
- 2) utilizar distribuciones estadísticas de enlace de hidrógeno y patrones de contactos no enlazados derivados de la Base de Datos Estructurales de Cambridge (Cambridge Crystallographic Data Centre, CB2 1EZ, U.K.).
- 3) emplear los resultados del programa GRID, para lo cual primero encuentra los mínimos y después selecciona un cierto número de puntos de baja energía alrededor de cada mínimo sobre los cuales fijar los fragmentos.

El siguiente paso consiste en ajustar los fragmentos moleculares sobre los sitios de interacción generados en el paso anterior. Para ello se explora la lista de sitios de interacción, utilizando criterios de distancia, para encontrar grupos adecuados de dos, tres o cuatro sitios sobre los que superponer los fragmentos moleculares: se calculan los cuadrados de las distancias entre los sitios de interacción y si estos valores se encuentran dentro de un determinado margen se procede a encajar el fragmento realizando un ajuste de mínimos cuadrados (RMS) entre los sitios diana definidos en el fragmento y los sitios de interacción, siempre y cuando el RMS sea menor de un valor umbral límite definido

por el usuario (generalmente entre 0.3 y 0.5 Å), y no haya ni solapamiento de van der Waals ni repulsión electrostática entre el fragmento y los átomos del receptor.

La fragmentoteca puede incluir también otras moléculas en conformaciones de baja energía obtenidas a partir de minimizaciones de mecánica molecular, y puede ser ampliada a voluntad. Los fragmentos se tratan como cuerpos rígidos pero se puede tener en cuenta la flexibilidad conformacional incluyendo en la fragmentoteca distintos conformeros para un mismo fragmento. Igualmente, como para fragmentos relativamente grandes, *e.g.* dipéptidos o compuestos heterocíclicos sustituidos, puede resultar difícil seleccionar los átomos que se van a utilizar en el proceso de superposición, es posible definir más de un conjunto de átomos diana por fragmento.

El último paso que realiza este programa es el de conectar los fragmentos ya encajados para dar lugar a una sola molécula. Para ello primero identifica los dos fragmentos más próximos entre sí y después aquellos hidrógenos de ambos fragmentos que a su vez se encuentran más cerca uno de otro. Estos dos hidrógenos, junto con los átomos a los que están unidos, se utilizan como sitios de referencia sobre los que situar el fragmento puente (*e.g.* -CH₂-CH₂-, -O-, -CO-, -CONH-, -SO₂NH-, etc), en un proceso análogo al anterior.

Uno de los inconvenientes de la actual implementación de LUDI es que, al considerar a la macromolécula de forma rígida, se rechazan fragmentos por no cumplir la distancia exigida, cuando ésta podría alcanzarse fácilmente mediante pequeños ajustes conformacionales de la cadena lateral de algún aminoácido clave. LUDI tampoco distingue entre enlaces de hidrógeno fuertes y débiles, no tiene en cuenta los efectos entrópicos, y requiere de un proceso de optimización posterior pues algunos enlaces y ángulos entre fragmentos y puentes pueden estar bastante alejados de sus posiciones de equilibrio. A pesar de estas limitaciones, y de tratarse esencialmente de un método geométrico, es una herramienta muy valiosa para el diseño de nuevas estructuras,

especialmente si se complementa con los mapas energéticos de GRID (Pisabarro, 1993), y compite muy ventajosamente en cuanto a rapidez con los cálculos de mecánica molecular sobre los diferentes complejos ligando-receptor.

11.0 CONSIDERACIONES FINALES.

El modelado molecular continúa su evolución aplicando una gran variedad de métodos computacionales al problema de identificar las complejas relaciones existentes entre estructuras moleculares y actividades biológicas en términos de interacciones entre los átomos constituyentes. El fin último de utilizar estas relaciones de forma predictiva para diseñar compuestos con un determinado perfil de actividad ya ha dado sus frutos en algunas ocasiones. Esta capacidad refleja la introducción tanto de nuevas metodologías como de ordenadores con la potencia suficiente para aplicar estos tratamientos al análisis de modelos moleculares detallados.

Aunque todas estas herramientas están proporcionando una valiosísima interpretación física de la actividad molecular, no hay que olvidar que la calidad de los resultados obtenidos va a depender de forma fundamental de la exactitud de las funciones energéticas utilizadas. Para el estudio de determinadas interacciones electrostáticas (Gallego, 1993) sería conveniente considerar, por ejemplo, las fluctuaciones en la polarización de las distribuciones electrónicas de átomos y moléculas, aspecto hasta ahora no cubierto por la mayor parte de los campos de fuerzas actuales, así como la dependencia de la distribución de cargas atómicas con la conformación.

La simulación microscópica de los acontecimientos que tienen lugar cuando dos moléculas difunden en disolución hasta encontrarse para producir un complejo no es todavía tratable computacionalmente debido al enorme número de moléculas de disolvente que serían precisas y a las distancias relativamente grandes que hay que salvar. Sin

embargo, estos procesos presentan mucho interés porque su influencia puede ser decisiva sobre la velocidad de la reacción. Para comprender cómo el comportamiento dinámico afecta a las constantes de velocidad de asociación se han desarrollado una serie de métodos de simulación molecular (Case, 1988) basados en técnicas de dinámica Browniana (McCammon, 1987b) que no se tratan en esta revisión no exhaustiva. Probablemente el ejemplo mejor estudiado es la difusión del anión superóxido al sitio activo de la superóxido dismutasa (Sharp, 1987; Case, 1988). Estas simulaciones no se limitan a la interacción de pequeñas moléculas con enzimas, sino que son aplicables al estudio de las asociaciones proteína-proteína y proteína-ADN.

Las técnicas computacionales pueden asistir tanto en el descubrimiento como en la optimización de compuestos "cabeza de serie" (Kuntz, 1992). En la práctica, está resultando más fácil descubrir nuevos compuestos activos con potencias del orden de micromolar que convertir estas moléculas en inhibidores con constantes de inhibición (K_i) del orden de nanomolar (Appelt, 1991). Una K_i baja, por otra parte, tampoco es garantía de selectividad absoluta sobre la diana estudiada, factor que adquiere especial relieve al considerar posibles efectos adversos. La ventaja de diseñar distintos cabezas de serie radica precisamente en que, aunque sus K_i sean parecidas, probablemente mostrarán diferentes perfiles farmacológicos y toxicológicos.

Los métodos de modelado molecular actualmente disponibles son posiblemente suficientes para tratar de solucionar multitud de problemas interesantes, aunque requieren buenas dosis de creatividad, ingenio y experiencia por parte del usuario. Cuanto menos, los modelos pueden ayudar a mejorar la percepción del problema bioquímico o farmacológico estudiado y pueden sugerir la realización de nuevos experimentos. En el mejor de los casos, proporcionan guías cualitativas, y en ocasiones cuantitativas, para el diseño de nuevos compuestos.

Las dificultades presentadas no hacen sino abrir nuevas vías de investigación que

brindan oportunidades inmejorables para aquéllos que desean dedicarse a la carrera científica.

12.0 BIBLIOGRAFÍA

1. Albert, A. (1979) "*Selective Toxicity*", 6th ed., Wiley, New York, p. 277.
2. Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A. et al. (1979) *Acta Cryst.* **B35**, 2331-2339.
3. Andrews, P. R.; Craik, D. J.; Martin, J. L. (1984) *J. Med. Chem.* **27**, 1648-1657.
4. Appelt, K.; et al. (1991) *J. Med. Chem.* **34**, 1925-1934.
5. Arnott, S.; Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun.* **47**, 1504-1510.
6. Badger, J.; Minor, I.; Kremer, M. J.; Oliveira, M. A. et al. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3304-3308.
7. Baniak, E. I.; Rivier, J. E.; Struthers, S.; Hagler, A. T.; Gierasch, L. M. (1987) *Biochemistry* **26**, 2642-2656.
8. Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. (1987a) *Science* **235**, 574-576.
9. Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. (1987b) *Science* **236**, 564-568.
10. Beddell, C.R.; Goodford, P.J.; Norrington, F.E.; Wilkinson, S.; Wootton, R. (1976) *Br. J. Pharm.* **57**, 201-209.
11. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F. Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. (1977) *J. Molec. Biol.* **112**, 535-542.
12. Besler, B. H.; Merz, K. M.; Kollman, P. A. (1990) *J. Comput. Chem.* **11**, 431-439.
13. Blake, C. C. F.; Koenig, D. G.; Mair, G. A.; North, A. C. T.; Phillips, D. C.;

- Sarma, V. R. (1965) *Nature* **206**, 757-761.
14. Blundell, T.; Sibanda, B. L.; Pearl, L. (1983) *Nature* **304**, 273-275.
 15. Bohacek, R.S.; McMartin, C. (1992) *J. Med. Chem.* **35**, 1671-1684.
 16. Böhm, H.-J. (1992a) *J. Comp.-Aided Mol. Des.* **6**, 61-78.
 17. Böhm, H.-J. (1992b) *J. Comp.-Aided Mol. Des.* **6**, 593-696.
 18. Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. (1982) *J. Biol. Chem.* **257**, 13650-13662.
 19. Boobbyer, D. N. A., Goodford, P. J., McWhinnie, P. M., Wade, R. C. (1989) *J. Med. Chem.* **32**, 1083-1094.
 20. Breslauer, K. J.; Remeta, D. P.; Chou, W.-Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J. G.; Marky, L. A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8922-8926.
 21. Bruccoleri, R. E.; Karplus, M. (1987) *Biopolymers* **26**, 137-168.
 22. Bruccoleri, R. E.; Karplus, M. (1990) *Biopolymers* **29**, 1847-1862.
 23. Brünger, A. T.; Karplus, M. (1991) *Acc. Chem. Res.* **24**, 54-61.
 24. Burt, C.; Huxley, P.; Richards, W. G. (1990) *J. Comput. Chem.* **11**, 1139-1146.
 25. Buttershell, C.; Malhotra, D.; Hopfinger, A. J. (1981) *J. Med. Chem.* **24**, 812-818.
 26. Case, D. A. (1988) *Prog. Biophys. Molec. Biol.* **52**, 39-70.
 27. Chang, G.; Guida, W. C.; Still, W. C. (1989) *J. Am. Chem. Soc.* **111**, 4379-4386.
 28. Charles, A. J.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C. (1991) *J. Am. Chem. Soc.* **113**, 1454-1455.
 29. Connolly, M. L. (1983a) *Science* **221**, 709-713.
 30. Connolly, M. L. (1983b) *J. Appl. Crystal.* **16**, 548-558.
 31. Cram, D. J. (1988) *Science* **240**, 760-767.
 32. Cramer, R.D. III; Patterson, D.E.; Bunce, J.D. (1988) *J. Am. Chem. Soc.* **110**, 5959-5967.

33. Crippen, G. M. (1977) *J. Comput. Phys.* **24**, 96-107.
34. Crippen, G. M. (1979) *J. Med. Chem.* **22**, 988-997.
35. Crippen, G. M. (1980) *J. Med. Chem.* **23**, 599-606.
36. Cushman, D.W.; Chueng, H.S.; Sabo, E.F.; Ondetti, M.A. (1977) *Biochemistry* **16**, 5484-5491.
37. Dauber-Osguthorpe, P.; Osguthorpe, D. J. (1990) *J. Am. Chem. Soc.* **112**, 7921-7935.
38. DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. (1988) *J. Med. Chem.* **31**, 722-729.
39. Dickerson, R. E.; Grzeskowiak, K.; Grzeskowiak, M.; Kopka, M. L.; Larsen, T.; Lipanov, A.; Privé, G. G.; Quintana, J.; Schultze, P.; Yanagi, K.; Yuan, H.; Yoon, H.-C. (1991) *Nucleosides Nucleotides* **10**, 3-24.
40. Donné-Op den Kelder, G. M. (1988) *Trends Pharmacol. Sci.* **9**, 391-393.
41. Gago, F.; Reynolds, C. A.; Richards, W. G. (1989) *Molec. Pharmacol.* **35**, 232-241.
42. Gago, F.; Richards, W. G. (1990) *Molec. Pharmacol.* **37**, 341-346.
43. Gallego, J.; Ortiz, A. R.; Gago, F. (1993) *J. Med. Chem.* **36**, 1548-1561.
44. Good, A. C.; So, S.-S.; Richards, W. G. (1993) *J. Med. Chem.* **36**, 433-438.
45. Goodford, P. J. (1985) *J. Med. Chem.* **28**, 849-857.
46. Goodsell, D. S.; Olson, A.J. (1990) *Proteins* **8**, 195-202.
47. Hagler, A. T.; Huler, E.; Lifson, S. (1974) *J. Am. Chem. Soc.* **96**, 5319-5327.
48. Hansch, C.; Leo, A. J. (1979) "*Substituent Constants for Correlation Analysis in Chemistry and Biology*", Wiley, New York
49. Hansch, C.; Klein, T.E. (1986) *Acc. Chem. Res.* **19**, 392-400.
50. Hibert, M. F.; Gittos, M. W.; Middlemiss, D. N.; Mir, A. K.; Fozard, J. R. (1988) *J. Med. Chem.* **31**, 1087-1093.

51. Hibert, M. F.; Hoffmann, R.; Miller, R. C.; Carr, A. A. (1990) *J. Med. Chem.* **33**, 1594-1600.
52. Hitzemann, R. (1988) *Trends Pharmacol. Sci.* **9**, 408-411.
53. Hopfinger, A. J. (1980) *J. Am. Chem. Soc.* **102**, 7196-7206.
54. Hopfinger, A. J. (1983) *J. Med. Chem.* **26**, 990-996.
55. Hopfinger, A. J. (1985) *J. Med. Chem.* **28**, 1133-1139.
56. Howard, A. E.; Kollman, P. A. (1988) *J. Med. Chem.* **31**, 1669-1675.
57. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. (1958) *Nature* **181**, 662-666.
58. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. (1983) *Science* **220**, 671-680.
59. Klebe, G.; Abraham, U. (1993) *J. Med. Chem.* **36**, 70-80.
60. Kline, A.D.; Braun, W.; Wüthrich, K. (1986) *J. Mol. Biol.* **189**, 377-382.
61. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. (1982) *J. Mol. Biol.* **161**, 269-288.
62. Kuntz, I. D. (1992) *Science*, **257**, 1078-1082.
63. Kuyper, L. F.; Roth, B.; Baccarani, D. P.; Ferone, R.; Beddell, C. R. et al (1985) *J. Med. Chem.* **28**, 303-311.
64. Labanowski, J.; Motoc, I.; Naylor, C.B.; Mayer, D.; Dammkoehler, R.A. (1986) *Quant. Struct.-Act. Relat.* **5**, 138-152.
65. Langridge, R.; Ferrin, T.E.; Kuntz, I.D.; Connolly, M.L. (1981) *Science* **211**, 661-666.
66. Lee, B.; Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379-400.
67. Lloyd, E.J.; Andrews, P.R. (1986) *J. Med. Chem.* **29**, 453-462.
68. Lolis, E.; Petsko, G. A. (1990) *Annu. Rev. Biochem.* **59**, 597-630.
69. Lumry, R.; Rajender, S. (1970) *Biopolymers* **9**, 1125-1227.
70. Mackay, D. H. J.; Cross, A. J.; Hagler, A. T. (1989) en "*Prediction of Protein*

- Structure and the Principles of Protein Conformation*", Fasman, G. D., ed. Plenum, New York, pp 317-358.
71. Manaut, M.; Sanz, F.; José, J.; Milesi, M. (1991) *J. Comput.-Aided Mol. Des.* **5**, 371-380.
 72. Marshall, G. R. (1985) *Ann. N.Y. Acad. Sci.* **439**, 162-169.
 73. Marshall, G. R.; Cramer III, R. D. (1988) *Trends Pharmacol. Sci.* **9**, 285-289.
 74. Marullo, S.; Emorine, L.; Strosberg, D.; Delavier-Klutcho, C. (1990) *EMBO J.* **9**, 1471-1476.
 75. Mayer, D.; Motoc, I.; Naylor, C. B.; Marshall, G. R. (1987) *J. Comp.-Aided Mol. Des.* **1**, 3-16.
 76. McCammon, J. A. (1987a) *Science* **238**, 486-491.
 77. McCammon, J. A.; Harvey, S. C. (1987b) "*Dynamics of Proteins and Nucleic Acids*", Cambridge University Press, Cambridge.
 78. Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. (1992) *J. Comput. Chem.* **13**, 505-524.
 79. Menziani, M. C.; De Benedetti, P. G.; Gago, F.; Richards, W. G. (1989) *J. Med. Chem.* **32**, 951-956.
 80. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H. (1953) *J. Chem. Phys.* **21**, 1087-1092.
 81. Morley, S. D.; Jackson, D. E.; Saunders, M. R.; Vinter, J. G. (1992) *J. Comput. Chem.* **13**, 693-703.
 82. Motoc, I.; Dammkoehler, R.A.; Mayer, D.; Labanowski, J. (1986) *Quant. Struct.-Act. Relat.* **5**, 99-105.
 83. Novotny, J., Brucoleri, R.; Karplus, M. (1984) *J. Mol. Biol.* **177**, 787-818.
 84. Novotny, J., Sharp, K. (1992) *Prog. Biophys. Molec. Biol.* **58**, 203-224.
 85. Oatley, S. J.; Blaney, J. M.; Langridge, R.; Kollman, P. A. (1984) *Biopolymers* **23**, 2931-2941.

86. Olson, A. J.; Goodsell, D. S. (1992) *Sci. Amer.* **11**, 44-51.
87. Orozco, M.; Luque, F. J. (1990) *J. Comput.-Aided Mol. Design* **4**, 411-426.
88. Ortiz, A. R., Pisabarro, M. T., Gallego, J., Gago, F. (1992), *Biochemistry* **31**, 2887-2896.
89. Ortiz, A. R.; Pisabarro, M. T.; Gago, F. (1993) *J. Med. Chem.* **36**, 1866-1879.
90. Page, M.I. (1977) *Angew. Chem Int. Ed. Engl.* **16**, 449-459.
91. Pattabiraman, N.; Levitt, M.; Ferrin, T. E.; Langridge, R. (1985) *J. Comput. Chem.* **6**, 432-436.
92. Pearlman, D. A.; Kollman, P. A. (1989) *Computer Simulations of Biomolecular Systems. Theoretical and Experimental Applications.* van Gunsteren, W. F.; Weiner, P. K., Eds. ESCOM, Leiden, 101-119.
93. Perutz, M. F.; Fermi, G.; Abraham, D. J.; Poyart, C.; Bursaux, E. (1986) *J. Am. Chem. Soc.* **108**, 1064-1078.
94. Pisabarro, M. T.; Ortiz, A. R.; Palomer, A.; Cabré, F.; García, L.; Wade, R.; Gago, F.; Mauleón, D.; Carganico, G. *J. Med. Chem.* en la imprenta.
95. Pullman, B.; Jortner, J., Eds. (1990) "*Molecular Basis of Specificity in Nucleic Acid-Drug Interactions*", Kluwer Academic Publishers, Dordrecht.
96. Quintana, J. R.; Lipanov, A. A.; Dickerson, R. E. (1991) *Biochemistry* **30**, 10294-10306.
97. Reynolds, C. A.; Richards, W. G.; Goodford, P. J. (1987) *Anti-Cancer Drug Design* **1**, 291-295.
98. Reynolds, C. A.; Essex, J. W.; Richards, W. G. (1992) *J. Am Chem. Soc.* **114**, 9075-9079.
99. Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.* **6**, 151-176.
100. Richards, W. G.; King, P. M.; Reynolds, C. A. (1989) *Protein. Eng.* **2**, 319-327.
101. Roberts, V. A.; Dauber-Osguthorpe, P.; Osguthorpe, D. J.; Levin, E.; Hagler, A. T.

- (1986) *Israel J. Chem.* **27**, 198-210.
102. Ryckaert, J. P.; Ciccoti, G.; Berendsen, H. J. C. (1977) *J. Comp. Phys.* **23**, 327-341.
 103. Sanz, F.; Manaut, F.; Rodríguez, J.; Lozoya, E.; López de Briñas, E. (1993) *J. Comput.-Aided Mol. Design* **7**, 337-347.
 104. Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. (1990) *J. Am. Chem. Soc.* **112**, 1419-1427.
 105. Schiffer, C. A.; Caldwell, J. W.; Kollman, P. A.; Stroud, R. M. (1990) *Proteins* **8**, 30-43.
 106. Sharp, K.; Fine, R.; Honig, B. (1987) *Science* **236**, 1460-1463.
 107. Sheridan, R. P.; Venkataraghavan, R. (1987) *Acc. Chem. Res.* **20**, 322-329.
 108. Shoichet, B. K.; Stroud, R. M.; Santi, D. V.; Kuntz, I. D.; Perry, K. M. (1993) *Science* **259**, 1445-1450.
 109. Tandford, C. (1980) "*The hydrophobic effect: formation of micelles and biological membranes*", 2nd ed., Wiley-Interscience, New York.
 110. Trumpp-Kallmeyer, S.; Hoflack, J.; Bruinvels, A.; Hibert, M. (1992) *J. Med. Chem.* **35**, 3448-3462.
 111. van Gunsteren, W. F. (1988), *Prot. Eng.* **2**, 5-13.
 112. Wade, R. C.; Clark, K. J.; Goodford, P. J. (1993a) *J. Med. Chem.* **36**, 140-147.
 113. Wade, R. C.; Goodford, P. J. (1993b) *J. Med. Chem.* **36**, 148-156.
 114. Ward, W. H. J.; Timms, D.; Fersht, A. R. (1990) *Trends Pharmacol. Sci.* **11**, 280-284.
 115. Watson, J. D.; Crick, F. H. C. (1953) *Nature* **171**, 964-967.
 116. Weiner, P. K.; Langridge, R.; Blaney, J. M.; Schaeffer, R.; Kollman, P. A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3754-3758.
 117. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. (1986) *J. Comput.*

Chem. **7**, 230-252.

118. Wender, P. A.; Koehler, K. F.; Sharkey, N. A.; Dell' Aquila, M. L.; Blumberg, P. M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4214-4218.
119. Wolfenden, R.; Radzicka, A. (1991) *Curr. Op. Struct. Biol.* **1**, 780-787.
120. Wong, C. F.; Mc Cammon, J. A. (1986) *J. Am. Chem. Soc.* **108**, 3830-3832.
121. Wüthrich, K. (1990) *Biochem. Pharmacol.* **40**, 55-62.