



# **Universitat Pompeu Fabra**

Departament de Ciències Experimentals i de la Salut

TESIS DOCTORAL

*Modelización molecular de los receptores de adenosina y sus  
ligandos en el marco de diseño de fármacos asistido por  
ordenador*

Memoria presentada por

**Hugo Gutiérrez de Terán Castañón**

Para optar al grado de Doctor

Barcelona, Febrero de 2004

El Doctorando

El Director de la Tesis Doctoral

**Hugo Gutiérrez de Terán Castañón**

**Prof. Ferran Sanz Carreras**

Dipòsit legal: B.43716-2004  
ISBN: 84-688-8524-X

*A mi madre, Magena  
A mi mujer, Raquel  
A mi hija, Lola*

**Invertir en ciencia es invertir en desarrollo**

Manifiesto por un Pacto de Estado por la Ciencia,  
propuesto por científicos españoles en febrero de 2004

## Agradecimientos

Esta es quizás la parte más difícil de escribir, pues ha sido tanta la gente que ha contribuido a que este momento llegase que es seguro que esta sección queda incompleta. Mis disculpas en ese caso.

Quiero agradecer vuestra participación, ayuda, comprensión o compañía a:

En primer lloc, al meu director de Tesi, mestre i amic, Prof. Ferran Sanz. Ha sigut un honor formar part del teu grup gaudint aquest anys de la teva experiència. A pesar de la teva agenda “terremoto” (mai havia vist una cosa igual), sempre hem trobat el moment de discutir i parlar de ciència i d’altres coses. Al Dr. Manolo Pastor, director de nuestro laboratorio y a quien considero codirector de esta tesis: sin tu ayuda el trabajo que aquí se recopila no sería el mismo. Al resto de miembros del GRIB (no os puedo nombrar a todos, ¡somos ya demasiados!); muy especialmente a Cristina Dezi y Fabien Fontaine (¡ánimo, sois los siguientes!), a quienes me une una gran amistad: que siga por muchos años; espero haberos ayudado casi tanto como vosotros a mí. Al Dr Jordi Rodrigo (¡Peich!) pues hemos compartido muchas cosas (menos eso que estás pensando, picarón) y de quien he aprendido mucho tanto en el terreno científico como personal. Mercé (de Jordi), no debo dejar de nombrarte aquí. Al Dr. Juanjo Lozano, nunca olvidaré tu inestimable ayuda en momentos que eran muy difíciles para ti. Dr. Jordi Villà, me has enseñado lo que no está escrito y me has ofrecido la oportunidad de trabajar contigo. Montse Barbany, lo que el MIPSim y Jordi han unido que no lo separe el tiempo. A Xavi Fustero, (qué grandes momentos en l’Escala), Alfons y Oscar (los Gonzalez), sin vosotros aun estaría tratando de configurar mi ordenador. A todo el laboratorio de genómica (Roderic-boys) mis pobres conocimientos de programación se deben a vosotros. Los de ping-pong, también. Dr. Baldo Oliva, Dr. Jordi Mestres, y su gente, habéis contribuido a que este grupo de investigación sea aún más interesante. Finalmente, a toda la sección administrativa (Maite, Raquel, Mireia, Carles, Eva, sin olvidar a Ester) pues sin vosotros esto no funcionaría. Merche y Yoli, entrar al IMIM por las mañanas sin veros no sería lo mismo. Yoli y Ferran, por amenizarnos estos años en Barcelona, aunque espero que la cosa no quede sólo ahí.

Special thanks to Prof. Johan Åqvist and all his group in Uppsala University. Those months on your lab. have been quite important for the development of the work here presented, but you also have been great hosts during my stay in Sweden. Ismael Tejero (Ismaelillo), con quien la experiencia “Gran Hermano” en Uppsala ha creado unos fuertes lazos de amistad. Pronto pasarás por este mismo trance ...

Gracias al Profesor Enrique Raviña y al Dr. Christian Fernandez Masaguer, por haberme brindado la oportunidad de trabajar en su laboratorio de Química Farmacéutica de la Universidad de Santiago de Compostela en lo que fue mi primera incursión en un trabajo de investigación. Este agradecimiento se hace extensivo a todos los compañeros de este laboratorio (Alberto, Bety, Eddi, Emilio, Yolanda) con quienes he compaginado trabajo, risas y amistad. A la Dra Mabel Loza (Farmacología, USC), pues siempre me ha tratado como a un colega y con gran cariño.

A todos mis amigos a los que he aburrido hasta la saciedad hablándoles de esta tesis que ahora concluye: Chente, Jose, Pelayo, Carlinos, Patri, Isma, Ruth, por nombrar a unos cuantos. Al recapitular me doy cuenta de lo importante que es teneros a mi lado. A Juaco (Dr. Díaz Alonso) y Gloria (Dra. Miranda), pues habéis sido los primeros en confiar en mi capacidad de dedicarme a la investigación.

A toda mi familia, pues me habéis dado los ánimos necesarios para continuar en esta carrera de obstáculos. Muy especialmente a mi madre, sin quien ni siquiera habría empezado y a quien hecho de menos enormemente en estos momentos. A mi hermano Iván: juntos continuamos quemando etapas (ahora vas tú) y has sido una pared maestra de este trabajo. A mi padre, a quien he de seguir intentando explicar de que va esto exactamente, pues se que confías en mi capacidad de trabajo, y a mi hermano Alfonso por interesarte en saber qué es lo que hago. A toda la familia de Raquel, que ya lo es mía también, (Carlos y Blanqui, y todos los demás habitantes de “La Casa del Patio”) pues me habéis apoyado y animado con vuestro optimismo.

Finalmente, a Raquel, mi mujer, amiga, amante y, en definitiva, compañera en la aventura de todos los días: sólo tú sabes lo importante que has sido en este trabajo que ahora concluye, yo no puedo ni siquiera intentar describirlo; y a Lola, mi hija, pues aunque de momento sólo puedo abrazarte y achucharte algún día sabrás lo que eso significa para mí.

## TABLA DE CONTENIDOS

<b>1. DISEÑO DE FÁRMACOS ASISTIDO POR ORDENADOR -----</b>	<b>9</b>
1.1. Métodos directos.....	11
1.1.1. Estructura experimental de proteínas .....	12
1.1.2. Modelización de proteínas .....	13
1.1.3. Predicción del modo de unión de ligandos .....	15
1.1.4. Evaluación de la interacción ligando-proteína .....	18
1.2. Métodos indirectos .....	21
1.2.1. QSAR.....	21
1.2.2. Farmacóforo, semejanza y diversidad moleculares.....	23
<b>2. METODOLOGÍA DE LA MODELIZACIÓN MOLECULAR -----</b>	<b>25</b>
2.1. Mecánica Cuántica .....	26
2.1.1. Propiedades derivadas de la función de onda molecular.....	27
2.2. Mecánica molecular .....	28
2.3. Dinámica molecular .....	32
2.4. Potenciales de interacción molecular.....	33
2.5. Métodos estadísticos .....	35
<b>3. ADENOSINA Y SUS RECEPTORES -----</b>	<b>39</b>
3.1. Adenosina: funciones, biosíntesis y degradación .....	39
3.2. Receptores acoplados a proteínas G .....	41
3.2.1. Activación y regulación de un GPCR .....	43
3.3. Clasificación de ARs.....	45
3.4. Interés terapéutico.....	46
3.5. Caracterización estructural.....	48
3.6. Agonistas y antagonistas .....	50
3.6.1. Agonistas.....	50
3.6.2. Antagonistas .....	52
<b>4. DISCUSIÓN -----</b>	<b>55</b>
4.1. Métodos directos (diseño de fármacos basado en receptor):.....	55
4.2. Métodos indirectos (diseño de fármacos basado en ligando):.....	58
<b>5. CONCLUSIONES -----</b>	<b>61</b>
<b>6. REFERENCIAS -----</b>	<b>63</b>
<b>7. ARTÍCULOS -----</b>	<b>75</b>



## 1. DISEÑO DE FÁRMACOS ASISTIDO POR ORDENADOR

Un *fármaco* es una molécula que interfiere en un determinado proceso bioquímico, bien sea fisiológico o fisiopatológico, de forma que dicho proceso se ve bloqueado o reforzado por la acción del fármaco. Esta acción se produce mayoritariamente por un mecanismo selectivo de interacción del fármaco con una macromolécula biológica (proteína, ADN, membrana celular); en este contexto diremos que el fármaco es un *ligando* de una *diana terapéutica*, la macromolécula biológica. Cuando el fármaco está incorporado en un vehículo apropiado para su administración en pacientes, de forma que puede alcanzar el proceso bioquímico que debe modular, entonces tenemos un *medicamento*, del cual decimos que el fármaco es el *principio activo*. Puede darse el caso, no obstante, de que un medicamento conste de una combinación de varios principios activos.

La estrategia actual de búsqueda de nuevos y mejores medicamentos es un proceso complejo e interdisciplinar que suele llevar de 10 a 12 años de investigación y desarrollo (I+D), con un gasto medio de 800 millones de €, tras lo cual tan solo un 5-10% de las moléculas que llegan a fase de ensayo clínico terminan siendo comercializadas.<sup>[1]</sup>

El uso de ordenadores, al igual que sucede en todas las áreas del conocimiento, ha facilitado y abaratado los costes de las diversas etapas de I+D de un nuevo medicamento. Aparte de las aplicaciones informáticas generales (comunicación, automatización de procesos, manejo de bases de datos ...) existen toda una serie de aplicaciones informáticas (tanto a nivel de *software* como de *hardware*) específicamente creadas para el sector biofarmacéutico. Así, la búsqueda de nuevas dianas terapéuticas emplea métodos bioinformáticos de predicción de genes,<sup>[2]</sup> procedimientos computacionales se pueden usar para filtrar bibliotecas moleculares o diseñar nuevos compuestos con alta probabilidad de presentar la actividad biológica deseada, las reacciones químicas pueden ser diseñadas o predichas mediante técnicas de química computacional,<sup>[3]</sup> la farmacocinética, el metabolismo y la toxicología se pueden predecir *in silico*<sup>[4,5]</sup> y existen diversas aplicaciones de informática médica para facilitar la gestión y el análisis de ensayos clínicos.<sup>[6]</sup> La presente tesis se centra en la utilización de métodos computacionales en el *diseño de nuevos ligandos* para una determinada diana terapéutica, lo que comúnmente se denomina “diseño de fármacos asistido por ordenador”.

El proceso de búsqueda de nuevos ligandos afines para una diana terapéutica arranca en el siglo XIX, siendo Paul Ehrlich quien introdujo el concepto de *receptor* (la macromolécula biológica que *recibe* al fármaco). Dichos ligandos se fueron descubriendo a partir del ensayo biológico de moléculas obtenidas a partir de fuentes naturales, como la codeína extraída de la planta *papaver somniferum*, o la penicilina que Fleming identificó en los hongos del género *penicilium*, o bien a partir de moléculas obtenidas por síntesis orgánica, como el LSD sintetizado y probado por Hoffman; más recientemente ciertos fármacos se sintetizaron racionalmente partiendo del conocimiento de la reacción bioquímica que se quería modular. Tal es el caso de la  $\alpha$ -metil-DOPA, que al ser un bloqueante de la enzima DOPA descarboxilasa actúa como fármaco antiparkinsoniano.

En los años 60 Hansch y Fujita plantearon relaciones cuantitativas entre parámetros fisicoquímicos moleculares y actividad biológica mediante técnicas estadísticas. El ordenador empezó a ser necesario para resolver estas relaciones, denominadas QSAR (*Quantitative Structure-Activity Relationships*).<sup>[7]</sup> Más adelante, en la década de los setenta, los métodos de cristalización de proteínas y resolución de su estructura mediante difracción de rayos X, así como los de resonancia magnética nuclear (RMN), empezaron a aportar suficientes estructuras de macromoléculas como para pensar en un diseño de

ligandos a partir del conocimiento tridimensional de su diana farmacológica. Durante la época de los 70 y 80 se produjo un gran desarrollo teórico de metodologías de modelización molecular y la aparición de los ordenadores personales permitió popularizar su aplicación. El penúltimo punto de inflexión en el desarrollo de fármacos tuvo lugar a mediados de los noventa, con la aparición de nuevas técnicas experimentales de química combinatoria, que permite una producción masiva de compuestos, y el cribado masivo (*High Throughput Screening*, HTS), para la consiguiente medición de sus actividades biológicas. El empleo de estas técnicas abrió las puertas al desarrollo de metodologías computacionales para el diseño racional de bibliotecas de compuestos, como paso previo a los experimentos de química combinatoria, y también para el cribado virtual *in silico* (*virtual screening*). Finalmente, la publicación del borrador del genoma humano<sup>[8]</sup> y de otros organismos supone una revolución en la biología, cuyas consecuencias aun está por valorar, aunque ya se empieza a hablar de era *post-genómica*, en la que la bioinformática, entendida como conjunto de aplicaciones informáticas para el tratamiento y procesado de información sobre biosecuencias, aparece como una destacada protagonista. Teniendo en cuenta que el proteoma es aun mayor que el genoma (debido a fenómenos de *splicing* alternativo y modificaciones post-transcripcionales), así como la variedad de dianas biológicas susceptibles de ser moduladas por un fármaco, (proteína, complejo proteína-proteína, ADN), se ha acuñado el término *druggable targetome*<sup>[9]</sup> para indicar la gran variedad de dianas susceptibles de ser atacadas por un fármaco, dando así a entender el amplio campo que queda abierto a partir del conocimiento del genoma humano.

Paralelamente, la otra revolución la constituye el desarrollo explosivo de la informática, que permite el procesado de volúmenes de datos enormes y la aplicación rutinaria, mediante cálculos de unas pocas horas, de complejas formulaciones teóricas como las de la química cuántica, que si bien eran conocidas hace años su resolución resultaba inabordable con los medios existentes. Actualmente, un laboratorio de fármacos asistido por ordenador cuenta con varios ordenadores personales con procesadores de alta velocidad, estaciones de trabajo y visualización y acceso a centros de supercomputación, propios o compartidos, que constan de *clusters* de decenas de procesadores que permiten paralelizar los cálculos.

En cuanto al abordaje de los problemas farmacológicos por técnicas computacionales, existen dos grandes formas de trabajar. Cuando disponemos de la estructura tridimensional de la diana terapéutica, bien obtenida por métodos experimentales (cristalografía de rayos X o RMN) o bien a través de la construcción de modelos moleculares, podemos abordar el diseño de ligandos basado en su estructura (*Structure based ligand design*), los llamados *métodos directos*. Si no disponemos de dichas estructuras, aun podemos utilizar éstas técnicas si obtenemos modelos por homología de suficiente confianza. En caso contrario, el diseño racional de ligandos se puede conseguir mediante *métodos indirectos*, basados en el análisis y comparación de propiedades moleculares y datos de afinidad por el receptor para ligandos conocidos, sin tener en cuenta la estructura de dicho receptor.<sup>[10]</sup> En la figura 1 se pretende dar una visión general de las técnicas utilizadas en una y otra metodología, las cuales se describirán en las secciones siguientes.

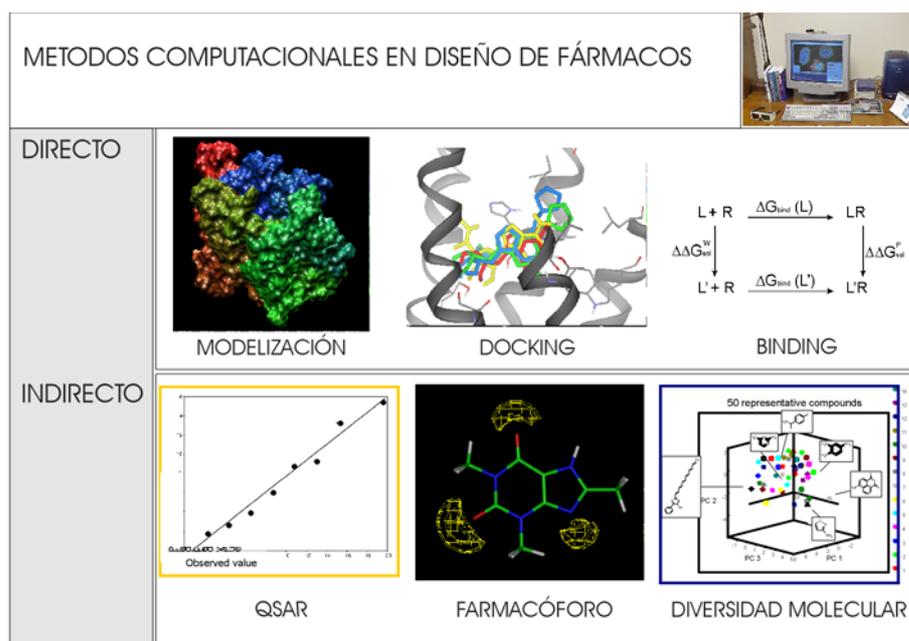


Figura 1: Métodos computacionales en el descubrimiento de nuevos fármacos:

### 1.1. Métodos directos

La posibilidad de diseñar fármacos partiendo de la estructura tridimensional de su diana terapéutica aparece con la resolución de las primeras estructuras de proteínas, la hemoglobina y la mioglobina,<sup>[11,12]</sup> por Max Perutz y John Kendrew, hecho que les valió el premio Nobel de Química de 1962. En 1977 se creó la base de datos de Brookhaven conocida como PDB (*protein data bank*), donde se han ido acumulando todas las estructuras experimentales publicadas de DNA y proteínas, alcanzándose 24.000 estructuras en enero de 2004. (<http://www.rcsb.org/pdb/>) El número de macromoléculas de estructura conocida es menor, ya que una proteína puede tener más de una entrada en el PDB. Para un número importante de macromoléculas se conoce la estructura co-cristalizada con un ligando, de forma que el sitio y modo de unión de éste sirven como un excelente punto de partida para diseñar nuevos fármacos. Existen también bases de datos que analizan específicamente las entradas correspondientes a estos complejos, como PDBsum (<http://www.biochem.ucl.ac.uk/bsm/pdbsum>) o Relibase. (<http://relibase.ebi.ac.uk/>) En los casos restantes, se acude a los experimentos de biología molecular y la simulación computacional para definir el sitio de unión de ligando (*binding site*).

En el caso de no disponer ni siquiera de la estructura de la proteína aislada, se puede obtener un modelo de ésta por métodos computacionales, sobre el cual trabajar de forma “directa” con los nuevos candidatos a fármaco.

Una vez que se dispone de una estructura de la proteína diana, bien sea a través de datos experimentales o de un modelo computacional de ésta, es necesario conocer dónde y cómo se unirán los ligandos a la proteína en cuestión.

Cuando se dispone de un modelo de interacción ligando-proteína fiable, se puede abordar la estimación de la afinidad del ligando por la proteína.

### 1.1.1. Estructura experimental de proteínas

Las técnicas experimentales que permiten la resolución de estructuras tridimensionales moleculares son las siguientes:

- **Cristalografía de rayos X (R-X):** Es la técnica que más estructuras macromoleculares ha reportado. La técnica se basa en obtener cristales del biopolímero, es decir, repeticiones ordenadas de unidades moleculares de éste, y someterlos a un bombardeo con rayos X. Debido a la corta longitud de onda de los rayos X (entre 0.2 y 2Å) los electrones de la molécula estudiada dispersan los rayos X, y las ondas dispersadas se recombinan de una forma que únicamente depende de la posición relativa de los átomos, la cual es exclusiva para cada molécula. El espectro de ondas refractadas es recogido por un detector para posteriormente, gracias a la transformación matemática conocida como transformada de Fourier, obtener un mapa de densidad electrónica del biopolímero, con una determinada resolución. A partir de este mapa se hacen las asignaciones de los núcleos, paso crucial y que implica subjetividad por parte del cristalógrafo. Dicha asignación se puede agilizar y mejorar utilizando programas computacionales *ad hoc*.

Un paso limitante en los estudios de R-X es la purificación y cristalización de la proteína. Si se trabaja con proteínas globulares, la disolución en medio acuoso facilita el proceso, mientras que para proteínas de membrana la cristalización es extremadamente complicada y existen muy pocas estructuras resueltas. En el caso concreto de una familia de receptores tan numerosa e importante como la de los GPCRs, hubo que esperar hasta el año 2000 cuando Palczewski y colaboradores obtuvieron la estructura de la rodopsina bovina,<sup>[13]</sup> con una resolución de 2.8 Å, valor que se considera de calidad intermedia.

Una limitación de las estructuras resueltas por rayos X es su rigidez, debido al empaquetamiento molecular en los cristales, así como la distorsión por la falta del entorno (acuoso en el caso de proteínas globulares, o lipídico en el caso de proteínas de membrana). Dicha información se puede obtener a través de técnicas de resonancia magnética nuclear.

- **Resonancia magnética nuclear (RMN):** En los años 80 Wüthrich y Ernst desarrollaron métodos para resolver estructuras proteicas por espectroscopía de RMN. El espectro de RMN recoge las frecuencias de un campo magnético externo, bajo las que los núcleos intrínsecamente magnéticos, como el  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  o  $^{31}\text{P}$ , absorben la energía necesaria para producir la transición entre los dos posibles momentos magnéticos de dicho núcleo. Esta frecuencia depende del núcleo que la absorbe, pero también del entorno en que dicho núcleo se encuentra. La influencia de los átomos directamente enlazados al núcleo que absorbe produce un “salto químico” (*chemical shift*) característico de cada grupo funcional. La aplicación de la RMN en la determinación de la estructura de proteínas es posible gracias al efecto NOE (Nuclear Overhauser Effect), producido por la interacción entre núcleos que no están covalentemente unidos y que es inversamente proporcional a la distancia que los separa, siempre que ésta sea menor a 5Å. Así se puede obtener, mediante espectros bidimensionales (llamados NOESY) información sobre la posición relativa de los núcleos en el espacio, identificándose pares de átomos separados por menos de 5Å. Combinando información de espectros RMN para dos o tres tipos de núcleos se obtienen espectros 3D-RMN ó 4D-RMN. La principal ventaja del RMN es su

aplicación sobre proteínas en disolución y la obtención de información de naturaleza dinámica.

- **Microscopía electrónica:** Esta técnica se emplea para obtener información sobre la organización supramolecular de las biomoléculas, debido a su capacidad para manejar macroestructuras de hasta 1500 Å de diámetro. A pesar de su baja resolución, su empleo ha sido de vital importancia en el estudio de proteínas de membrana. Los experimentos con rodopsina<sup>[14,15]</sup> fueron, durante años, la única forma de conocer la orientación y agrupamiento de las siete hélices transmembrana de la rodopsina,<sup>[16]</sup> la cual sirvió de patrón para obtener los primeros modelos de receptores acoplados a proteína G (GPCRs).

### 1.1.2. Modelización de proteínas

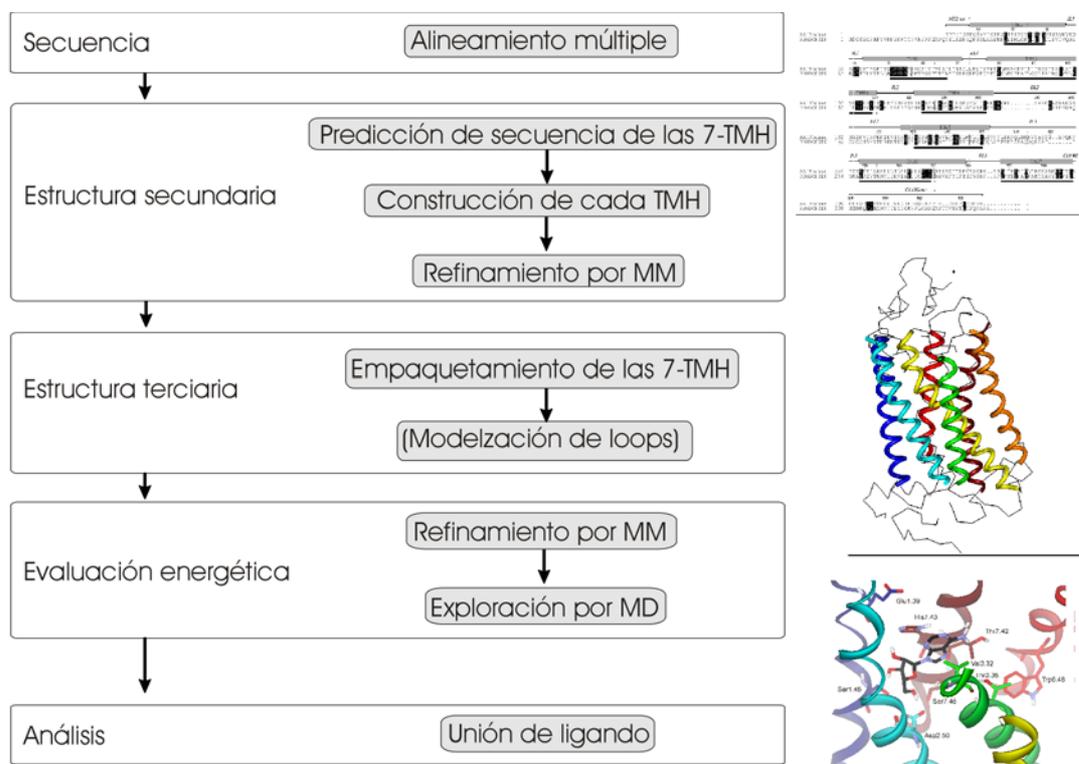
Cuando no se dispone de la estructura experimental de la proteína problema, aún se pueden utilizar métodos directos de diseño de ligandos acudiendo a técnicas de modelización molecular.<sup>[17]</sup> Con éstas se pretende obtener un modelo teórico de la estructura de la proteína, que si bien en ningún caso será una representación exacta, nos puede permitir explorar las zonas de interés de la proteína con una cierta confianza.

Las técnicas de modelización de proteínas se pueden clasificar de la siguiente forma<sup>[18,19]</sup>:

- **Métodos *ab initio*:** estos métodos intentan predecir la conformación nativa de una secuencia partiendo únicamente de su secuencia<sup>[20]</sup>. La idea subyacente es que toda la información necesaria para plegar una secuencia está en la propia secuencia.<sup>[21]</sup> Es necesaria una representación simplificada de la estructura proteica a estudiar, realizar una exploración exhaustiva del espacio conformacional y disponer de un método que permita una evaluación energética rápida de las múltiples geometrías generadas. Un ejemplo de estos métodos es el programa ROSETA.<sup>[22]</sup>
- **Reconocimiento de plegamiento (*fold recognition*):** la idea de esta técnica es buscar cuáles de los plegamientos proteicos conocidos pueden ser similares al plegamiento desconocido de la proteína problema, de la que sólo conocemos la secuencia.<sup>[23]</sup> La observación de que proteínas no relacionadas entre sí (a nivel de secuencia) adoptan a menudo plegamientos similares, se ha intentado explicar desde distintos puntos de vista, lo que ha llevado al desarrollo de distintos algoritmos de predicción de estructura terciaria. Así, existen métodos que trabajan más sobre secuencia primaria, como el programa FFAS,<sup>[24]</sup> otros, por el contrario, se fijan sobre todo en las leyes físicas que gobiernan el proceso de plegamiento<sup>[25]</sup>. Por último, una combinación de ambos criterios da lugar a las llamadas técnicas de *threading*, como en el programa THREADER.<sup>[26]</sup>
- **Modelado por homología:** estas técnicas son aplicables cuando la proteína problema (*target*) es similar a una de estructura tridimensional conocida (proteína patrón o *template*). Se considera que la identidad de secuencia entre ambas proteínas para abordar un modelado por homología ha de ser al menos de un 30%.<sup>[27]</sup> El proceso implica una serie de pasos, en cada uno de los cuales el investigador ha de tomar decisiones, lo cual justifica su presencia y seguimiento del proceso. No obstante, hay paquetes de programas que automatizan bastante el proceso, como es el caso de MODELLER.<sup>[28]</sup> Un caso extremo de obtención de modelos automáticos es el servidor SWISS-MODEL (<http://www.expasy.org/swissmod/SWISS-MODEL.html>), que únicamente necesita que el investigador envíe la secuencia y en pocos minutos le

devuelve un modelo de la proteína. Sin embargo esta automatización, si bien puede ser útil para generar ideas, no es recomendable para obtener modelos de calidad.<sup>[29]</sup>

- Modelado por topología de proteínas transmembrana:** Bajo este epígrafe se comentarán los protocolos utilizados para modelar las proteínas que atraviesan la membrana celular y que no tienen un patrón homólogo lo suficientemente cercano (identidad de secuencia < 30%), y por tanto no son susceptibles de modelar por homología. Es necesario por tanto identificar los segmentos transmembrana y predecir la orientación global de éstos en la membrana.<sup>[30]</sup> Un protocolo específico para la familia de los GPCRs ha sido propuesto por Ballesteros y Weinstein en 1994,<sup>[31]</sup> tomando entonces como topología patrón el mapa de proyección de la densidad electrónica de rodopsina bovina sobre la membrana.<sup>[15]</sup> Este protocolo se esquematiza en la figura 2.



**Figura 2:** Esquema del método de modelización de GPCRs propuesto por Ballesteros y Weinstein. Abreviaturas: TMH (*hélices transmembrana*), MM (*mecánica molecular*), MD (*dinámica molecular*). Las imágenes corresponden a la modelización del receptor de adenosina A<sub>1</sub>, descrito en el Artículo II

Existen ejemplos previos en nuestro grupo de investigación de la aplicación de este esquema para la construcción de modelos de GPCRs, los cuales han servido para explicar sucesos bioquímicos tales como el paso inicial de la activación del receptor adrenérgico del tipo  $\alpha_{1d}$  (en colaboración con la Universidad de Bari)<sup>[32]</sup>, o la unión diferencial de agonistas y antagonistas a los receptores de serotonina 5-HT<sub>2A</sub> y 5-HT<sub>2C</sub>. No obstante, con la obtención de la estructura de la rodopsina a 2.8Å de resolución<sup>[13]</sup>, muchos grupos han optado por aplicar un protocolo de modelado por homología utilizando esta estructura

como patrón. Sin embargo, dada la baja homología entre la familia de GPCRs ( $< 20\%$ ), existe el riesgo de sobrerrepresentar la información de la secuencia patrón (rodopsina) en la estructura final de nuestra secuencia problema, sobre todo en la zona de los loops. De todas formas, la información derivada de la estructura de Palczewski ha mejorado sustancialmente la calidad de los modelos de GPCRs. Debido a la falta de información experimental sobre la estructura de estas proteínas, la consideración de los datos de experimentos de mutagénesis dirigida o construcción de receptores quiméricos es de vital importancia para asesorar la validez de los modelos así construidos. También es necesaria una exploración exhaustiva del acoplamiento de alguno de los ligandos conocidos para la GPCR problema, y cotejar el modelo obtenido con los datos experimentales disponibles, ya que las mayores diferencias estructurales entre la familia de GPCRs se dan en la zona de unión de ligandos, y es ahí donde la información de la estructura de rodopsina, por sí sola, será insuficiente.

### 1.1.3. Predicción del modo de unión de ligandos

Predecir cómo se unirá un nuevo ligando a la proteína de interés no es en ningún caso una tarea trivial ni libre de subjetividad. Dependiendo del tipo de información experimental de que dispongamos, el problema se puede abordar de distintas formas. Si la proteína es relativamente fácil de cristalizar, y sobre todo si tiene un alto interés farmacológico, acostumbran a existir complejos ligando-receptor cristalizados. En estos sistemas se puede disponer de información “privilegiada”, como es la posición relativa de distintos ligandos en el sitio activo de la proteína, información que se puede procesar por métodos computacionales para ayudar a alinear nuevos ligandos y así conocer así su posición de acoplamiento en la proteína, como se ha mostrado en un estudio reciente de nuestro laboratorio de investigación.<sup>[33]</sup> En los casos en los que únicamente se disponga de uno de estos complejos cristalizados, ya hay información suficiente para conocer el *sitio* de unión de los nuevos fármacos. Quedará entonces por deducir el *modo* de unión de estas nuevas moléculas. Si además tenemos datos de biología molecular que nos dan pistas sobre las interacciones que han de tener lugar entre fármaco-proteína (por ejemplo, experimentos de mutagénesis dirigida), puede ser abordable un acoplamiento manual entre ambas moléculas, con la ayuda de programas de modelización molecular. Este tipo de ensayos busca la mayor concordancia entre datos experimentales y modelos moleculares, obteniéndose modelos que explican todo lo posible los datos experimentales.

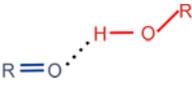
Por el contrario, existen muchos métodos automáticos para explorar los posibles modos de unión de nuevos ligandos a la proteína problema. Son los llamados programas de *docking* (anglicismo que se puede traducir por acoplamiento), que realizan una exploración exhaustiva de posiciones relativas ligando-proteína, evaluando las interacciones intermoleculares en cada posición explorada. Como resultado de esta exploración se obtiene una colección de posibles posiciones de acoplamiento ligando-proteína, que el programa ordena según el valor que su función de evaluación le ha dado a cada solución (*scoring*). Estos dos pasos consecutivos (exploración y evaluación) se han resuelto de muy diversas formas en los programas de *docking* disponibles. A continuación comentaré brevemente las distintas estrategias utilizadas en cada uno de estos pasos:

- **Exploración geométrica:** idealmente, a la hora de buscar modos de unión ligando-receptor habría que considerar todas las combinaciones posibles de grados de libertad: traslación ( $x, y, z$ ), rotación ( $\phi, \theta, \psi$ ), flexibilidad de ligando (número de enlaces

rotables) y flexibilidad de proteína ( $\frac{\text{n}^\circ \text{ de enlaces rotables}}{\text{residuo}} \times \text{n}^\circ \text{ residuos}$ ). En la práctica, ningún algoritmo de búsqueda puede abordar un problema tan complejo en un tiempo de cálculo razonable. Así, los primeros programas desarrollados hacían un docking totalmente rígido, sin considerar flexibilidad de proteína ni de ligando (DOCK3, AUTODOCK2.4, GROUP); el siguiente paso fue implementar la flexibilidad del ligando (DOCK4.0, AUTODOCK3.0, FLEX-X). Finalmente la flexibilidad de la proteína se ha incorporado en algoritmos como QXP,<sup>[34]</sup> aunque de manera parcial. Se han ideado distintas estrategias indirectas para abordar este último problema de forma más amplia,<sup>[35]</sup> como el uso en simulaciones paralelas de distintos conformeros de la proteína, la utilización de librerías de rotámeros de cadenas laterales o resumir las distintas conformaciones de la proteína en una única geometría difusa, como se ha hecho recientemente con FlexX<sup>[36]</sup> o AUTODOCK.<sup>[37]</sup> Se han implementado una gran variedad de algoritmos de búsqueda: *Monte Carlo simulated annealing* (QXP, AUTODOCK2.4), construcción incremental del ligando en la cavidad de la proteína (DOCK, FlexX), algoritmo genético (GOLD) o algoritmo genético lamarkiano (AUTODOCK3.0).

- **Evaluación energética:** El propósito de una función de evaluación (*scoring function*) es discriminar, en un tiempo razonable de cálculo, entre las soluciones *correctas* (posiciones con un bajo rmsd -desviación estandar- con respecto a la posición cristalográfica) y el resto de soluciones encontradas en el proceso de búsqueda. Con este objetivo se hace una estimación de la energía de unión del complejo ligando-receptor en cada posición. La interacción entre dos moléculas, como pueden ser un fármaco y su receptor, viene dada por una suma de fuerzas intermoleculares de carácter no enlazante. Estas fuerzas aparecen esquematizadas en la tabla 1, y su naturaleza se irá comentando con detalle en las siguientes secciones. La energía de interacción fármaco-receptor resultante de todas estas interacciones se puede estimar de diferentes maneras:<sup>[38,39]</sup>

**Tabla 1:** Interacciones intermoleculares entre fármaco y receptor.

<i>Tipo de interacción</i>	<i>Geometría óptima</i>	<i>Ejemplo</i>
Electrostática o iónica	distancia: 2.8 Å	$=\text{NH}_2^+ \cdots \text{OOC}^-$
Puente de hidrógeno	distancia: 2.7-3.1 Å ángulo: 120-180°	
Van der Waals	distancia: 3-4 Å	$\text{CH}_3 \cdots \text{H}_3\text{C}^-$
Cambios entrópicos	-	Interacción hidrofóbica Energía conformacional

- Basándose en campos de fuerza: se puede estimar la energía de unión ligando-receptor según los principios físicos que subyacen en los campos de fuerza moleculares (véase Sección 2.2). Los algoritmos de docking pueden

incorporar el campo de fuerzas completo (como hace el módulo de docking GROUP del programa GRID<sup>[40]</sup>) o versiones simplificadas de estos campos de fuerza, lo que ayuda a acelerar el proceso de cálculo. Así, por ejemplo, QXP y AUTODOCK utilizan versiones simplificadas del campo de fuerza de AMBER.<sup>[41]</sup> No obstante, AUTODOCK3.0 modifica este campo de fuerzas mediante la introducción de parámetros empíricos (véase Sección 2.4)

- Potenciales estadísticos: se puede asignar un valor para cada tipo de interacción no enlazante entre dos moléculas (ver tabla 1) a través de unos parámetros empíricos, y posteriormente sumar los valores correspondientes a todas las interacciones detectadas y obtener así una estimación empírica de la energía de unión. Hablamos entonces de funciones empíricas de evaluación (*empirical scoring functions*), como la implementada en FlexX<sup>[42]</sup>:

$$\Delta G_{bind} = \Delta G_0 + \Delta G_{rot} \times N_{rot} + \Delta G_{Hb} \sum_{Hb} f(\Delta R, \Delta \alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta \alpha) + \Delta G_{aro} \sum_{aro} f(\Delta R, \Delta \alpha) + \Delta G_{lipo} \sum_{lipo} f^*(\Delta R) \quad (1)$$

donde cada término  $\Delta G$  es un valor estimado para cada tipo de interacción (Hb = puente de hidrógeno, aro = aromáticas, lipo = lipofílicas, ionic = iónicas), y  $f(\Delta R, \Delta \alpha)$  es un factor inversamente proporcional a la desviación sobre la geometría óptima de la interacción;  $f^*(\Delta R)$  se refiere al número de contactos hidrofóbicos y  $N_{rot}$  es el número de enlaces rotables.

Otro tipo de potenciales estadísticos son las llamadas “funciones basadas en el conocimiento” (*knowledge-based scoring functions*), derivadas de las estructuras de complejos ligando-proteína conocidas experimentalmente a través de la mecánica estadística. Tal es el caso de los potenciales de fuerza media (PMF, *potentials of mean force*) de Muegge y Martín.<sup>[43]</sup> En esta aproximación, los átomos se clasifican en tipos de átomo (tipo  $i$  para los átomos de ligando, tipo  $j$  para los de la proteína) y se que suman todas las interacciones que se dan entre átomos del ligando ( $k$ ) y átomos de la proteína ( $l$ ), siempre que éstos se encuentren, en el complejo a estudiar, a una distancia  $r$  menor que la distancia de corte dada para ese par  $i, j$  de tipos de átomo,  $r_{cut-off}$ :

$$E_{PMF} = \sum_{r^{kl} < r_{cut-off}^{ij}} -k_B T \ln \left[ f_{Vol\_corr}^j(r) \frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}} \right] \quad (2)$$

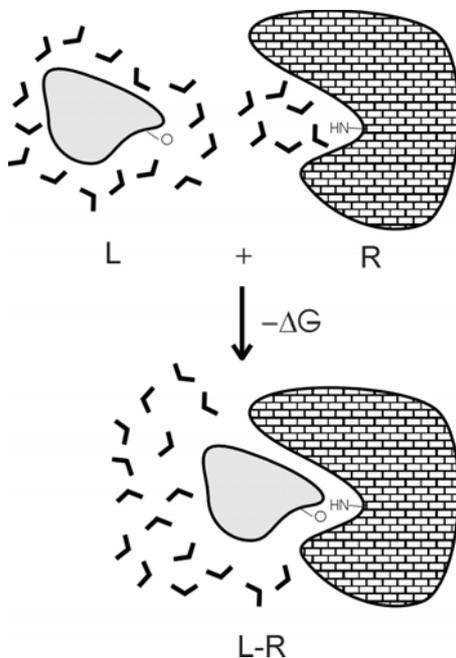
siendo  $K_B$  la constante de Boltzman,  $T$  es la temperatura,  $f_{Vol\_corr}^j(r)$  es un factor de corrección relacionado con el volumen del ligando, y  $\frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}}$  es la relación entre el número de veces que el par  $i, j$  aparece interaccionando en la base de datos de complejos experimentales a la distancia  $r$  y el número de veces que dicho par aparece sin interaccionar. La energía de interacción ligando-receptor se estima, corrigiendo  $E_{PMF}$  a través de factor empírico  $\varepsilon$ :

$$\Delta G_{bind} = E_{PMF} / \varepsilon \quad (3)$$

Existen otras *knowledge-based scoring function*, que incorporan otros parámetros, como la llamada DrugScore.<sup>[44]</sup>

### 1.1.4. Evaluación de la interacción ligando-proteína

En el apartado anterior se han mostrado algunos métodos rápidos de evaluación de la interacción ligando-receptor en el marco de una exploración exhaustiva del acoplamiento. En este apartado hablaré de los métodos que existen para evaluar de una forma más sofisticada la energía de interacción del ligando con el receptor. Estos métodos implican una mayor demanda computacional, y por tanto no pueden aplicarse de forma rutinaria en procesos masivos de cribado virtual y deben reservarse para los casos en los que se ha identificado un modo “fiable” de unión para el ligando.



**Figura 3:**

Esquema del proceso de unión ligando (L) y receptor (R), para formar el complejo ligando-receptor (L-R), con la consiguiente liberación de energía (signo negativo del término  $\Delta G$ ).

¿Qué es lo que sucede cuando un ligando se une a su receptor? El proceso está esquematizado en la figura 3, donde un ligando y su receptor están inicialmente separados, solvatados ambos por moléculas de agua. Tras la unión (*binding*), ambas moléculas aparecen estabilizadas por interacciones intermoleculares no enlazantes, y el reordenamiento de las moléculas de agua, así como la libertad conformacional del ligando, han cambiado. Desde el punto de vista termodinámico existe un cambio de energía libre entre las dos situaciones, expresado mediante la ecuación:

$$\Delta G_{binding} = \Delta H - T\Delta S \quad (4)$$

La contribución entrópica a la energía libre ( $\Delta S$ ) viene dada por la pérdida de libertad conformacional del ligando (reducción de entropía) y por el desordenamiento de las moléculas de agua que formaban estructuras ordenadas alrededor de ligando y receptor, al asociarse éstos y reducir su superficie de acceso a solvente (efecto hidrofóbico, que aumenta la entropía).

Al término entálpico ( $\Delta H$ ) contribuyen las interacciones moleculares de tipo no enlazante entre ligando y receptor (ver tabla 1).

Podremos relacionar la  $\Delta G_{\text{binding}}$  con medidas experimentales como la constante de disociación de un complejo,  $K_d$ :

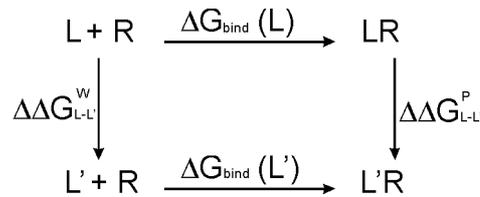
$$K_d = \frac{[R][L]}{[RL]} \quad (5)$$

mediante la ecuación:

$$K_d = \exp\left(\frac{\Delta G_{\text{binding}}}{RT}\right) \quad (6)$$

Como se ha visto anteriormente, los métodos de *docking* incorporan funciones que evalúan de forma rápida el valor de  $\Delta G_{\text{binding}}$  utilizando aproximaciones de mecánica molecular o potenciales estadísticos.

No obstante, también es posible calcular la diferencia de energía libre entre dos procesos, por ejemplo entre la unión de dos ligandos distintos a una misma proteína, según el siguiente ciclo termodinámico:



De donde se deduce que la diferencia de energía libre entre los dos acoplamientos ligando-receptor (energía relativa de *binding*) se puede calcular, teóricamente, como la diferencia entre cualquiera de los dos pares de ramas paralelas del ciclo:

$$\Delta \Delta G_{\text{bind}}^{(L \rightarrow L')} = \Delta G_{\text{bind}}(L) - \Delta G_{\text{bind}}(L') = \Delta \Delta G_{L \rightarrow L'}^P - \Delta \Delta G_{L \rightarrow L'}^W \quad (7)$$

Las ramas horizontales no son calculables en la práctica; sin embargo, cada una de las ramas verticales (la transformación de un ligando  $L$  en otro ligando  $L'$  en medios acuoso y en el entorno de la proteína) se puede calcular por el método conocido como perturbación de energía libre (*free energy perturbation*, FEP).<sup>[45]</sup> La aplicación de esta técnica requiere ir transformando un sistema en otro gradualmente, a lo largo de un parámetro de conversión  $\lambda$  que toma valores  $i$  que van de 0 (potencial de  $L$ ) a 1 (potencial de  $L'$ ) en una serie de  $n$  pasos. En cada uno de estos pasos se calcula la diferencia de energía ( $\Delta G_i$ ) para pasar del estado  $i$  al estado  $i+1$  mediante la fórmula (8)

$$\Delta G_i = -k_B T \cdot \ln \left\langle \exp \left( -\frac{1}{k_B T} (V_{(i+1)} - V_{(i)}) \right) \right\rangle_{(i)} \quad (8)$$

Donde  $V_i$  es el potencial de “*mapping*” correspondiente a  $V_i = (1-\lambda_i)V_L + \lambda_i V_{L'}$ . Esto implica obtener valores promedio (indicados por los corchetes de la ecuación) de las diferencias entre las energías potenciales ( $V_{(i+1)}$  y  $V_{(i)}$ ) calculados sobre conformaciones generadas en el potencial de una de ellas ( $i$ ), obtenidas mediante muestreo de dinámica molecular o análisis de Monte Carlo. Finalmente se suman las diferencias de energía

obtenidas en cada paso para obtener la diferencia energética entre ambas moléculas en el entorno considerado (agua o proteína):

$$\begin{array}{ccc}
 L & \xrightarrow[0 \quad \lambda \quad 1]{\frac{\Delta G_i}{0} \quad \frac{\Delta G_i}{\lambda} \quad \frac{\Delta G_i}{1} \quad \frac{\Delta G_i}{1}} & L' \\
 \Delta G_{L \rightarrow L'} = \sum_{i=1}^n \Delta G_{i \rightarrow i+1} & & (9)
 \end{array}$$

En la práctica los dos potenciales  $V_L$  y  $V_{L'}$  no han de ser muy distintos entre sí, para llegar a la convergencia.

No obstante su perfección teórica, este método presenta grandes inconvenientes, como son: su elevado coste computacional, problemas de convergencia de los valores calculados de los potenciales ( $V$ ), la necesidad de calcular estos valores para especies químicas intermedias inexistentes en la realidad y la necesidad de que los ligandos  $L$  y  $L'$  sean muy parecidos en su estructura. Por estas razones se han desarrollado otros métodos de cálculo de energía de *binding* que se sitúan entre las sencillas funciones de *scoring* y los complejos cálculos de FEP, incluyendo distintas aproximaciones teóricas para poder calcular la  $\Delta G_{binding}$  considerando únicamente los extremos del ciclo termodinámico. Así, el grupo de Kollman ha desarrollado el método MM-PBSA,<sup>[46]</sup> en el que es necesario estimar la energía del ligando libre, del receptor libre y del complejo ligando-receptor:

$$\Delta G_{bind} = \Delta G_{complejo} - (\Delta G_{ligando} + \Delta G_{receptor}) \quad (10)$$

Otro método interesante, utilizado en el Artículo III de esta tesis, es el método de la energía de interacción lineal (*linear interaction energy*, LIE), desarrollado por Åqvist y colaboradores,<sup>[47]</sup> en el que únicamente es necesario considerar los estados físicamente relevantes del ligando: libre (solvatado en agua) y asociado (en el complejo ligando-receptor). El cambio de la energía libre entre uno y otro estado corresponderá a la energía libre de unión. En éste método, la energía de interacción del ligando con el entorno se descompone en dos términos: un término electrostático ( $V_{l-s}^{el}$ ) y un término no electrostático, ( $V_{l-s}^{vdW}$ ) el cual tiene en cuenta interacciones de tipo van der Waals e hidrofóbicas. De esta forma es posible el empleo de la aproximación de la respuesta lineal (LRA, *Linear Response Approximation*) para interacciones electrostáticas,<sup>[47,48]</sup> lo que conduce a la expresión:

$$\Delta G_{el} = \frac{1}{2} \left( \langle V_{l-s}^{el} \rangle_p - \langle V_{l-s}^{el} \rangle_w \right) = \frac{1}{2} \left( \Delta \langle V_{l-s}^{el} \rangle \right) \quad (11)$$

en la que la diferencia de la energía de interacción electrostática del ligando con el medio en sus dos estados (libre y asociado con la proteína) aparece ponderada por un factor de  $\frac{1}{2}$ .

De forma análoga, la diferencia en la energía de interacción no electrostática entre ambos estados podría aproximarse como la diferencia entre los términos de van der Waals ponderada por un factor empírico,  $\alpha$ :

$$\Delta G_{vdW} = \alpha \left( \langle V_{l-s}^{vdW} \rangle_p - \langle V_{l-s}^{vdW} \rangle_w \right) = \alpha \left( \Delta \langle V_{l-s}^{vdW} \rangle \right) \quad (12)$$

La energía libre total de unión corresponde a la suma de las contribuciones electrostática y no electrostática, lo que conduce a la expresión:<sup>[49]</sup>

$$\Delta G_{bind} = \alpha \Delta \langle V_{l-s}^{vdW} \rangle + \beta \Delta \langle V_{l-s}^{el} \rangle + \gamma \quad (13)$$

Como se ha explicado, el parámetro  $\beta$  toma valor 0.5 si se asume la aproximación de respuesta lineal; sin embargo, estudios posteriores muestran desviaciones de esta aproximación, en particular para ligandos neutros dipolares y para aquellos que contienen grupos hidroxilo,<sup>[50]</sup> obteniéndose valores del coeficiente  $\beta < 0.5$  para estos ligandos. La constante  $\gamma \neq 0$  es un término empírico que es necesario introducir en algunos sistemas, a fin de ajustar los valores absolutos de las  $\Delta G_{\text{binding}}$  calculadas.

En el Artículo III y su discusión se profundiza sobre el uso de este método en el sistema biológico de interés en esta tesis, los receptores de adenosina.

## 1.2. Métodos indirectos

Cuando se desconoce – o no se desea tener en cuenta – la estructura del receptor que constituye la diana farmacológica, el análisis de un conjunto de ligandos nos puede dar información muy útil para diseñar moléculas más potentes sobre dicho receptor.

### 1.2.1. QSAR

Es posible relacionar las variaciones estructurales en una serie de ligandos con variaciones en la actividad de forma cuantitativa, a través de las llamadas ecuaciones de relaciones estructura actividad o QSAR (*Quantitative structure-activity relationships*)<sup>[7]</sup> Ésta idea fue ya formulada por Richet a finales del siglo XIX, a través de la ecuación:

$$\Delta\Phi = f(\Delta C) \quad (14)$$

Siendo C la estructura química y  $\Phi$  la medida de actividad.

No fue hasta 1964 cuando aparecieron las primeras aproximaciones QSAR que se aplicaron con éxito al diseño de nuevas moléculas. Dos métodos fueron desarrollados en paralelo, que abordaban el tema desde planteamientos teóricos diferentes:

- Aproximación de **Hansch**:<sup>[51]</sup> también llamada aproximación extratermodinámica, supone que la energía libre de unión ligando-receptor se puede aproximar mediante una combinación lineal de contribuciones lipofílica, electrónica y estérica:

$$\Delta G = \Delta G_{\text{lipo}} + \Delta G_{\text{el}} + \Delta G_{\text{est}} \quad (15)$$

Así se propuso relacionar la actividad (expresada en función de la concentración de producto necesaria para una determinada actividad, C) con una combinación lineal de parámetros físico-químicos, relativos a la lipofilia (por ejemplo, el coeficiente de partición entre una fase hidrofóbica y otra hidrofílica, generalmente octanol y agua, denominado logP) y a la electronegatividad (como el parámetro de Hammet,  $\sigma$ ),<sup>[52]</sup> a través de ecuaciones del tipo:

$$\log \frac{1}{C} = k_1 \log P + k_2 \sigma + k_3 \quad (16)$$

Donde las constantes  $k_i$  se obtienen mediante métodos de regresión sobre los datos de una serie congénica cuyas actividades se conocen. Posteriormente se añadieron modificaciones, como la incorporación de parámetros estéricos o la formulación de relaciones parabólicas con respecto a la lipofilia.

- Aproximación de **Free-Wilson**.<sup>[53]</sup> En este caso se trataba de considerar las aportaciones que hacen a la actividad de un compuesto sus distintos sustituyentes químicos ( $\chi_i$ ) localizados en posiciones  $j$  de la estructura  $\mu$ :

$$\log 1/C = \sum \chi_{ij} + \mu \quad (17)$$

Aunque inicialmente el método fue pensado para construir moléculas *de novo*, a partir de la adición de los distintos fragmentos, únicamente se llegó a aplicar en estudios de relaciones estructura-actividad de series congenericas.

Cuando las relaciones QSAR hacen uso de descriptores fisico-químicos que dependen de la estructura tridimensional de la molécula, entonces se habla de *3D-QSAR*.<sup>[7]</sup> Su uso está muy extendido actualmente en la investigación farmacéutica, no sólo para predecir las actividades sino también para estimar otras propiedades (toxicológicas, farmacocinéticas). Dos herramientas son necesarias para un estudio 3D-QSAR: calcular los descriptores químicos tridimensionales y relacionarlos con actividades a través de análisis estadístico multivariante.

- Descriptores: en una relación 3D-QSAR interesa describir propiedades químicas de una molécula que tengan que ver con su posible *binding*. Estas propiedades estarán relacionadas con las fuerzas de interacción molecular resumidas en la tabla 1. Debido a que estas interacciones ligando-receptor son óptimas a una determinada distancia  $r$ , es útil construir una rejilla alrededor de cada ligando y calcular en cada punto de esta rejilla su potencial de interacción molecular (MIP, *Molecular Interaction Potential*) con determinados grupos químicos, llamados sondas (*probes*). En el caso de que el cálculo del MIP considere una sonda protón, y dicho cálculo considere la función de onda de la molécula, entonces se habla de potencial electrostático molecular (MEP *Molecular Electrostatic Potential*). A la distribución de valores de MIP alrededor de una molécula se la denomina campo de interacción molecular (MIF, *molecular interaction field*). En la Sección 2.4 se explica el proceso de obtención de los diferentes tipos de potenciales de interacción molecular.
- Análisis estadístico: una vez así calculados los MIFs de una serie de moléculas, podemos construir una matriz que relacione cada valor de actividad con los valores del MIF. Esta matriz tendrá tantas filas como moléculas ( $n$ ) y tantas columnas como puntos tenga el MIF más el valor de actividad ( $x + 1$ ). La resolución de este sistema viene de la aplicación del método de análisis multivariante PLS (*partial least squares*, mínimos cuadrados parciales), comentado brevemente en la Sección Sección 2.5.

Los métodos 3D-QSAR más ampliamente utilizados son el método CoMFA (*Comparative Molecular Field Analysis*)<sup>[54]</sup> y el método GRID/GOLPE.<sup>[55]</sup> La principal diferencia entre estos métodos es el tipo de descriptores que usan: el método CoMFA calcula dos tipos de campos de interacción molecular: estérico (un potencial Lennard-Jones 12-6) y electrostático (potencial culómbico); el método GRID/GOLPE permite calcular un campo para cada sonda química implementada en el programa GRID<sup>[40]</sup>, que utiliza aproximaciones de la mecánica molecular (ver Sección 2.4). También hay diferencia en el tratamiento estadístico ya que, si bien los dos métodos utilizan PLS, en el GRID/GOLPE se han implementado una serie de pre-tratamientos de los valores de los MIFs.<sup>[56,57]</sup>

Un paso crítico en los métodos 3D-QSAR es el alineamiento de las moléculas teniendo en cuenta los grupos responsables de una interacción común con el receptor, o *grupos farmacofóricos*. Este paso puede hacerse de forma manual, en el caso de trabajar con un conjunto pequeño de compuestos, aunque es reconocida la dificultad y elevado gasto de tiempo en este proceso, o bien de forma automática. Para ésta última opción se han desarrollado programas que trabajan una superposición automática sobre puntos farmacofóricos, como DISCO, ALADDIN o CATALYST o el módulo t-fit del programa QXP.<sup>[34]</sup> Pero, ¿qué sucede cuando no se han determinado unas zonas comunes de interacción con el receptor, o bien el alineamiento de las moléculas no es evidente?

### 1.2.2. Farmacóforo, semejanza y diversidad moleculares

Si se han determinado los grupos farmacofóricos necesarios para la interacción con el receptor de cualquier potencial ligando, a este grupo se le denomina *farmacóforo*. Si se establecen relaciones geométricas entre los distintos grupos farmacofóricos, entonces tenemos un farmacóforo 3D.<sup>[58]</sup> Éste concepto es tremendamente útil, ya que la contraparte de este farmacóforo estará formado por los grupos del receptor que interaccionan con los grupos farmacofóricos, con lo cual podemos analizar la distribución espacial de estos grupos del receptor y obtener una imagen esquemática del sitio de unión. Como se ha comentado en el apartado anterior, existen varios programas computacionales dedicados a la búsqueda de farmacóforos 3D, que superponen las moléculas y buscan sus conformaciones para dicha superposición. Una estrategia original que intenta identificar relaciones geométricas entre puntos farmacofóricos sin necesitar de un alineamiento molecular está implementada en el programa ALMOND.<sup>[59]</sup> Este método se basa en codificar la información de los MIF, obtenidos con el programa GRID, en unos descriptores independientes de alineamiento (Grid Independent Descriptors, GRIND). Ello se consigue seleccionando un número determinado de nodos representativos del MIF, para posteriormente codificar las interdistancias entre cada par de nodos en un espectro, denominado correlograma. Los correlogramas obtenidos para cada molécula se pueden superponer, lo cual permite su uso para buscar diferencias en éstos que se puedan relacionar con la actividad, en un estudio 3D-QSAR, o puntos comunes de este correlograma, en una búsqueda de farmacóforo. Debido a la posibilidad de una transformación inversa de correlograma a nodos de MIF, este método es fácilmente interpretable en el marco de estudios de diseño de nuevos fármacos. El método es aplicable también al estudio de sitios de unión en macromoléculas, para describir lo que antes he denominado “contraparte del farmacóforo”. En el Artículo II se describe el uso de este método para la caracterización farmacofórica de sitios de unión a ribosa.

Para hablar rigurosamente de farmacóforo es necesario incluir en su definición moléculas con actividad sobre una cierta diana farmacológica y que a la vez sean químicamente diversas. Estas moléculas presentan a menudo un alineamiento (o superposición) no evidente, correspondiente a la superposición óptima de sus campos de interacción molecular, ya que es la disposición de estos campos la que guiará su superposición en el sitio de unión.

Para abordar este problema, en nuestro grupo de investigación se ha desarrollado el programa MIPSim (Molecular Interaction Potentials Similarity).<sup>[60]</sup> El módulo COMP de este programa compara pares de moléculas sobre la base de sus distribuciones tridimensionales de MIP calculadas en una rejilla de puntos definida alrededor de las moléculas. El MIP puede ser obtenido mediante interacción con sondas del programa GRID, o bien ser el MEP obtenido a partir del cálculo de la función de onda (ver Sección

2.4). La medida de semejanza se obtiene a través de un coeficiente, generalmente el coeficiente gaussiano,<sup>[60,61]</sup> aunque también es posible utilizar otros coeficientes como el de Spearman o el de Hodking.<sup>[62]</sup> El programa también permite, a través del módulo MIN, calcular relaciones geométricas entre los puntos de máxima interacción con las sondas, en un nuevo intento de definición de farmacóforo 3D. Un ejemplo reciente de la aplicación del programa MIPSim ha sido la aplicación al estudio de análogos de estado de transición de reacciones enzimáticas, para lo cual se han utilizado comparaciones del MEP entre el estado de transición y análogos propuestos en la literatura de la enzima corismato mutasa. (véase apéndice IV y ref.<sup>[63]</sup>)

Actualmente esta metodología se está utilizando en nuestro laboratorio en el estudio de antagonistas de receptores de adenosina, basándonos en resultados previos obtenidos a partir de la comparación de tres antagonistas clásicos. (ver apéndice III) Por otro lado, es posible obtener alineamientos biológicos de ligandos de un determinado receptor para un estudio 3D-QSAR, siempre que dispongamos de datos estructurales suficientes, mediante la optimización de parámetros que ponderen la combinación de los coeficientes de semejanza obtenidos con varios MIFs. (véase apéndice V y ref.<sup>[33]</sup>)

Otro aspecto interesante de la similaridad molecular en el diseño de fármacos es su opuesto, es decir, el análisis de la *diversidad* molecular. El desarrollo de las técnicas de química combinatoria, que permiten la síntesis en paralelo de miles de compuestos, hace necesaria una selección racional de los sustituyentes introducidos en una determinada posición, de forma que cubran la máxima variabilidad de propiedades moleculares con una selección de una fracción de los componentes disponibles en una base de datos. Para ello se describen todos los sustituyentes posibles mediante descriptores moleculares y se lleva a cabo un análisis matemático con el fin de elegir un subconjunto de  $n$  sustituyentes que presenten la máxima diversidad y representatividad sobre el espacio de descriptores. El tipo de descriptores químicos utilizados varían desde valores unidimensionales (peso molecular), bidimensionales (como los índices de conectividad) o tridimensionales (como los descriptores derivados del MIF). Por otro lado, el manejo de datos de un volumen tan grande de compuestos, al igual que en el 3D-QSAR, se hace mediante técnicas de análisis multivariante, en este caso análisis de componentes principales, seguido de agrupamiento de compuestos en conglomerados. (*cluster analysis*) Un ejemplo de este tipo de estudio se detalla en el Artículo I. Posteriormente, en nuestro laboratorio se ha hecho un estudio comparativo sobre la utilización de distintos tipos de descriptores químicos en el estudio de una base de datos de reactivos. (véase apéndice II y ref.<sup>[64]</sup>)

## 2. METODOLOGÍA DE LA MODELIZACIÓN MOLECULAR

Existen dos grandes áreas en la química computacional dedicadas al desarrollo de modelos moleculares, a fin de explicar la estructura de las moléculas y su reactividad: las técnicas de **Mecánica Molecular** y las de la **Química Cuántica**. Dentro de cada área se han desarrollado infinidad de técnicas, y en este apartado únicamente pretendo dar una visión general de la metodología existente haciendo especial hincapié en las técnicas que se han empleado a lo largo de la presente tesis.

Todos los métodos, independientemente de su fundamento teórico, pretenden abordar los siguientes problemas:

- Calcular la energía asociada a una estructura molecular determinada y a partir de ésta poder derivar una serie de propiedades asociadas con ella. Las propiedades que se puedan calcular dependerán de los fundamentos teóricos del método empleado. Por ejemplo, los métodos cuánticos permiten calcular calores de formación de forma bastante precisa, mientras que mecánica molecular permite la simulación de la dinámica molecular, con lo que es posible predecir energías libres unión ligando-proteína.
- Optimizar la geometría de un sistema molecular, o en otras palabras, localizar la estructura molecular con menor energía (por lo que se utilizan indistintamente los términos optimización geométrica y minimización energética). Esta optimización se puede realizar a través del cálculo de las derivadas de la energía con respecto a las grados de libertad geométricos (obteniendo los métodos basados en el *gradiente de energía*), buscando desplazar el sistema en una dirección que conduzca a un valor menor de energía. Los métodos basados en el gradiente presentan el inconveniente de que tienden a conducir el sistema hacia mínimos de energía próximos a la posición de partida (*mínimos locales*) que no tienen por qué coincidir con el mínimo absoluto que corresponde a la geometría óptima que se pretende encontrar. Algunos métodos habituales basados en el gradiente son:
  - *Steepest descent*: Resulta útil como primera aproximación, en sistemas que están alejados del mínimo local más cercano. Mediante cálculos del gradiente, se evalúan los cambios en la energía del sistema asociados a perturbaciones sobre los grados de libertad geométricos, y se modifica el sistema en la dirección indicada por el gradiente. El proceso se repite hasta que el cambio energético generado por la perturbación de cualquier grado de libertad geométrico es menor que un umbral predefinido.
  - *Conjugated gradients*: A diferencia del método anterior, tras cada evaluación de gradiente se tiene en cuenta la información obtenida de evaluar el gradiente en dicho paso para el paso posterior. Ello supone un mayor coste de cálculo, por lo que este método se emplea para explorar geometrías que están cerca de un mínimo de energía.
  - Existen métodos todavía más costosos computacionalmente, como el *Newton-Raphson*, que incorporan el cálculo de la segunda derivada de la energía con respecto a las posiciones atómicas, con la finalidad de obtener información sobre la velocidad de cambio en la función gradiente y modular adecuadamente la evolución del sistema hacia la geometría óptima.
- Simular la variación a lo largo del tiempo de una determinada estructura molecular (dinámica molecular). Aunque tradicionalmente este estudio era exclusivo de la

mecánica molecular, actualmente se empiezan a abordar para sistemas pequeños a través de técnicas cuánticas.<sup>[65]</sup> No obstante, en esta tesis únicamente se discutirán dinámicas moleculares basadas en los postulados de la mecánica molecular.

## 2.1. Mecánica Cuántica

La Química Cuántica aborda la distribución de electrones en un sistema molecular, codificada mediante la función de onda molecular,  $\Psi$ . Ésta se relaciona con la energía del sistema mediante la ecuación de Schrödinger, que si consideramos su formulación independiente del tiempo, toma la forma de la siguiente ecuación diferencial:

$$H\Psi_{(r)} = E\Psi_{(r)} \quad (18)$$

siendo  $H$  una función diferencial que incluye la energía cinética y potencial de núcleos y electrones, denominada operador hamiltoniano. Esta ecuación es una “ecuación de valor propio” (*eigenvalue*), debido a que el operador,  $H$ , actuando sobre la función,  $\Psi$ , produce un múltiplo de la función, en este caso multiplicada por el valor de la energía,  $E$ . La resolución de esta ecuación nos llevará a varias soluciones posibles, correspondientes a estados estacionarios de la molécula, cada uno caracterizado por una determinada función de onda  $\Psi$  con un valor de energía  $E$  asociado a ella. La solución que obtiene el menor valor de energía corresponde al estado basal (*ground state*). Dicha resolución es compleja, y requiere la introducción de una serie de aproximaciones:

- Aproximación de Born-Openheimer: asume que el movimiento de los núcleos se puede despreciar, debido a que es infinitamente más lento que el de los electrones.
- Combinación lineal de orbitales atómicos: según el “producto de Hartree”, la función de onda del sistema se puede expresar como un producto de  $n$  orbitales moleculares monoeléctricos:

$$\Psi(r) = \phi_1(r_1)\phi_1(r_1)\dots\phi_n(r_n) \quad (19)$$

en el que cada orbital molecular monoeléctrico  $\phi$  se expresa, de acuerdo con la aproximación LCAO (*Linear Combination of Atomic Orbitals*), como una combinación lineal de  $N$  orbitales atómicos ( $\chi$ ), también llamados funciones de base:

$$\phi_i = \sum_{\mu=1}^N c_{\mu i} \chi_{\mu} \quad (20)$$

A su vez, cada función de base se puede expresar de forma análoga como una combinación lineal de una serie de funciones matemáticas de tipo gaussiano. La calidad de la aproximación dependerá del conjunto de funciones utilizado para dichas combinaciones lineales (llamado *basis set*). Actualmente es común utilizar una función de base 6-31G\* para calcular propiedades de moléculas orgánicas pequeñas (tipo fármaco). 6-31G\* indica que se utilizan seis funciones gaussianas para representar los orbitales atómicos internos, mientras que los externos –capa de valencia– se representan mediante tres funciones para la parte “contraída” y una para la parte “difusa”, más una función de polarización para átomos pesados (indicada mediante el \*), que representa el efecto de los orbitales tipo  $d$ . Existen, no obstante, otras funciones de base menos costosas de calcular, que se pueden utilizar para la búsqueda de la geometría óptima, como son las bases STO-3G o 3-21G.

Debido a que cada electrón se halla bajo el campo eléctrico del resto, cada orbital molecular monoeléctrico dependerá de los demás, con lo que la búsqueda del mejor conjunto de coeficientes se hará de forma iterativa. Esta aproximación recibe el nombre de campo autoconistente. (*Self-Consistent Field*, SCF). Dentro de los métodos que utilizan cálculos SCF cabe destacar el método Hartree-Fock (HF), que es el de mayor difusión en el diseño de fármacos y en esta tesis ha sido el único método utilizado de los mencionados anteriormente. Las denominadas ecuaciones de Hartree-Fock buscan encontrar la función de onda que produzca el mínimo de energía de un sistema molecular, es decir, imponen la condición de que la primera derivada de la energía sea cero ( $\delta E=0$ ) sujeta a su vez a la condición de que los orbitales moleculares sean ortonormales (teorema variacional). Este método se puede mejorar teniendo en cuenta la interacción de configuraciones (CI, *configuration interaction*), aunque con un elevado coste de cálculo. Finalmente, existen otras aproximaciones, como la basada en la teoría funcional de la densidad (DFT, *density functional theory*).

Todos estos métodos, debido a que consideran toda la estructura atómica y no introducen parámetros empíricos en la resolución de la ecuación de Schrödinger, se denominan métodos *ab-initio*. En cuanto a los programas con que se realizan cálculos químico-cuánticos *ab-initio*, entre los más comunes están el paquete GAUSSIAN<sup>[66]</sup> o GAMESS<sup>[67]</sup>

Cuando se introducen parámetros ajustados empíricamente que sustituyen alguna de las integrales de los métodos *ab-initio*, pero aún consideramos la estructura electrónica de la molécula, nos hallamos ante métodos semiempíricos. Estos métodos omiten los cálculos relativos a los electrones de capa interna, (*core*) que tratan como si formasen parte del núcleo y se centran únicamente en los electrones de capas de valencia, partiendo de la base de que son éstos electrones los responsables de la reactividad y la mayoría de propiedades moleculares. No obstante, no existen parámetros validados para todos los sistemas moleculares, lo cual hace aconsejar prudencia en su utilización. Se han desarrollado multitud de métodos semiempíricos, entre los cuales destacan, por lo común de su utilización en el campo de desarrollo de fármacos, el MNDO<sup>[68]</sup> o el AM1<sup>[69]</sup>, implementados ambos en el popular programa MOPAC.<sup>[70]</sup> Es común referirse a estas parametrizaciones con el nombre “hamiltoniano” seguido de las siglas del método (por ejemplo, hamiltoniano AM1). Existe también la posibilidad de incorporar el efecto del solvente en los cálculos semiempíricos, considerando a éste como un sistema continuo de un dieléctrico determinado, mientras que el soluto se trata a través de un hamiltoniano semiempírico. A este respecto, los grupos de Cramer y Truhlar (U. De Minnesota) han introducido varios modelos de solvatación, tanto para agua como para otros solventes orgánicos, en el programa AMSOL.<sup>[71]</sup> En estos modelos el solvente es descrito a través de unos parámetros (índice de refracción, constante dieléctrica, acidez, basicidad...), y permiten escoger entre varios hamiltonianos semiempíricos para el soluto. Uno de estos modelos (SM5.4/AM1) ha sido utilizado para calcular diversas propiedades moleculares en el Artículo I. Para una revisión de estos modelos véase la bibliografía recogida en la dirección web: <http://t1.chem.umn.edu/amsol/smxref.htm>

### 2.1.1. Propiedades derivadas de la función de onda molecular

Procesos como los de optimización pueden ser calculadas para moléculas pequeñas tipo fármaco, de varias decenas de átomos, en un tiempo razonable de cálculo a través de métodos mecánico-cuánticos. Explorar las moléculas pequeñas considerando su

distribución electrónica puede tener varias ventajas con respecto a la utilización de la mecánica molecular, a pesar del mayor coste computacional:

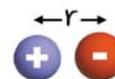
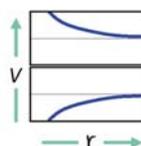
- Los métodos de mecánica molecular utilizan una serie de parámetros empíricos, los cuales han sido derivados para macromoléculas biológicas. Por ello, es frecuente encontrarse con que alguno de los parámetros necesarios para caracterizar moléculas pequeñas, tipo fármaco, no esté incluido en el campo de fuerzas. Como se verá más adelante, algunos de los métodos de mecánica molecular obtienen parámetros desconocidos, como el valor de las cargas atómicas o el de las distancias y ángulos de equilibrio, mediante cálculos químico-cuánticos.
- Ciertos procesos, como la formación y ruptura de enlaces, sólo son abordables considerando la distribución electrónica de las moléculas implicadas. Estos procesos son cruciales, por ejemplo, en el estudio de reacciones enzimáticas. Para poder considerar a la vez los efectos cuánticos de la reacción enzimática y el entorno de la proteína (sólo abordable mediante métodos de mecánica molecular), se han desarrollado métodos híbridos QM/MM (*quantum mechanics / molecular mechanics*).<sup>[72-75]</sup> La caracterización de estados de transición de una reacción química por métodos químico-cuánticos permite la búsqueda de análogos de dichos estados de transición, los cuales son un excelente punto de partida para el diseño de nuevos inhibidores enzimáticos<sup>[76]</sup> o para la búsqueda de nuevos catalizadores de una reacción química.<sup>[63]</sup>
- De la función de onda se pueden derivar multitud de descriptores moleculares, que pueden ser utilizados en la caracterización de moléculas. Entre estos cabe citar:
  - Potencial electrostático molecular (MEP): Esta propiedad es de gran interés en la modelización molecular ya que describe las características electrónicas de las moléculas y se puede utilizar en el análisis y predicción de interacciones moleculares.<sup>[17]</sup> En la Sección 2.4 explico su relevancia y métodos de obtención.
  - Entalpía de formación ( $\Delta H_f$ ): derivada del cálculo de la energía, da cuenta de la estabilidad de un compuesto. Puede ser calculada en distintos entornos (agua, solventes orgánicos) en los métodos que consideran medios distintos del vacío.
  - Momento dipolar ( $\mu$ ): es una magnitud vectorial relacionada con la asimetría de la distribución de carga de un compuesto; da idea de su polaridad.
  - Energía de los orbitales moleculares: tienen especial interés la del orbital ocupado de mayor energía (HOMO, *highest occupied molecular orbital*) y la del desocupado de menor energía (LUMO, *lowest unoccupied molecular orbital*). Ambos están implicados en interacciones intermoleculares de transferencia de carga.

## 2.2. Mecánica molecular

Las técnicas de mecánica molecular consideran una simplificación de la molécula como un sistema de núcleos atómicos conectados por unos enlaces; la simplificación es a menudo comparada como un sistema de bolas y muelles. Esta analogía responde a las siguientes características:

- Cada bola encierra las propiedades de un átomo, propiedades que como se ha explicado dependen en realidad de su distribución electrónica. En la mecánica molecular, en lugar de calcular la función de onda en cada sistema, se estima el valor de las propiedades derivadas de ésta a través de una serie de parámetros. La primera consideración es definir un conjunto de átomos posibles (*atom types*), cada uno de los cuales viene determinado no sólo por el elemento a que se refiere (oxígeno, carbono ...) sino por los diferentes estados en que éste se puede encontrar en un sistema molecular (carbono con hibridación sp<sup>2</sup>, carbono con hibridación sp<sup>3</sup> ...). Una vez definido el conjunto de *atom types*, el efecto de la distribución electrónica se puede aproximar a través de la inclusión de una serie de cargas parciales (*partial charges*) para cada átomo. Estas cargas se habrán de definir para cada molécula con la que se trabaja, puesto que dependen no solo del *atom type* sino del entorno molecular. El método de cálculo de carga parcial varía desde aproximaciones topológicas, como las cargas de Gasteiger-Marsili,<sup>[77]</sup> a métodos teóricamente más exactos, como el *restrained electrostatic potential fitting* (RESP)<sup>[78]</sup> empleado para definir las cargas parciales en el campo de fuerzas de AMBER.<sup>[41]</sup> Este método obtiene las cargas parciales mediante un ajuste que pretende reproducir lo mejor posible la distribución del potencial electrostático molecular, calculada ab initio con una base 6-31G\*. Existen métodos intermedios, como los basados en el análisis de Mulliken de la distribución electrónica sobre el hamiltoniano AM1. Es común el empleo de los métodos de cálculo de cargas más rápidos (Gasteiger, AM1) en los estudios de docking, mientras que procedimientos más elaborados como el RESP se reservan para estudios de dinámica molecular o para el cálculo de energía libre de unión. Las cargas parciales son necesarias para calcular interacciones de tipo electrostático que se dan entre dos átomos *i,j* que no están enlazados, a través del término de interacción culómbica:

$$V_{el(i,j)} = -\frac{1}{4\pi\epsilon} \frac{q_i q_j}{r_{ij}} \quad (21)$$



fórmula

representación gráfica

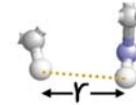
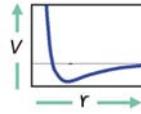
esquema

Siendo  $q$  la carga parcial de cada átomo y  $r$  la distancia que los separa;  $\epsilon$  la constante dieléctrica del medio, que representa la oposición del entorno (agua y proteína) a la interacción entre ambas cargas. El valor de esta constante es de  $\epsilon \cong 80$  para el agua, mientras que los valores típicos de proteínas son mucho menores (entre 2 y 5), si bien pueden existir microentornos proteicos más polares donde el valor de  $\epsilon$  se incrementa.<sup>[79]</sup> Se puede afirmar que el uso de esta constante pretende incorporar todos aquellos efectos que no han sido explícitamente considerados en el modelo (como el solvente acuoso o la membrana celular), tomando valor unidad cuando todos estos efectos se consideran de manera explícita. La optimización de su valor en cálculos referidos a proteínas es motivo de continuas revisiones,<sup>[80]</sup> si bien una aproximación generalizada es asignar valores de  $\epsilon$  dependientes de la distancia  $r_{ij}$ .<sup>[75,81]</sup>

Otros parámetros que es necesario definir son las constantes  $A$  y  $B$  que corresponden respectivamente a los términos de repulsión y atracción de la función

de van der Waals, la cual evalúa la interacción no electrostática entre pares de átomos no enlazados:

$$V_{vdW(i,j)} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \quad (22)$$



fórmula

representación gráfica

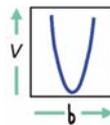
esquema

Estos valores se obtienen empíricamente para cada *atom type* interaccionando consigo mismo ( $A_{ii}$  y  $B_{ii}$ ). Los valores para cada par  $ij$  se obtienen según mediante la media geométrica  $A_{ij} = \sqrt{A_{ii} \cdot A_{jj}}$  (y de forma análoga para  $B$ ). Existen formas alternativas de formular la función de van der Waals, en base a variaciones en el valor de los exponentes de los términos  $r_{ij}$ . La formulación de la ecuación (22) recibe el nombre de potencial de Lennard-Jones 12-6.

Los términos electrostático ( $V_{el}$ ) y de van der Waals ( $V_{vdW}$ ) se calculan para cada par de átomos no enlazados que se encuentren a una distancia  $r$  menor que una distancia de corte  $r_{\text{cut-off}}$ . Esta aproximación se utiliza para ahorrar tiempo de cálculo, ya que si no habría que evaluar  $N^2$  interacciones (siendo  $N$  el número de átomos del sistema). Existen otras aproximaciones para evaluar la interacción electrostática a un  $r > r_{\text{cut-off}}$ , como el método *local reaction field*.<sup>[48]</sup>

- Se asume que los enlaces entre átomos son flexibles alrededor de unas distancias de equilibrio ( $r_0$ ), de forma que pueden vibrar como lo hace un muelle. Esta vibración se representa mediante un potencial harmónico cuadrático del tipo:

$$V_{\text{enlace}} = k_r (r - r_0)^2 \quad (23)$$



fórmula

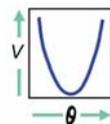
representación gráfica

esquema

Donde  $k_r$  es una constante de fuerza definida para cada par de *atom-types* ( $ij$ ) que puedan estar enlazados, y  $r_0$  la distancia de equilibrio del enlace  $ij$ .

- De forma análoga, el ángulo  $\theta$  que forma cada par de enlaces contiguo también pueden variar alrededor de un valor de equilibrio  $\theta_0$ , siendo la constante de fuerza  $k_\theta$ :

$$V_{\text{ángulo}} = k_\theta (\theta - \theta_0)^2 \quad (24)$$



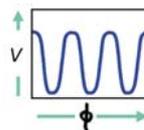
fórmula

representación gráfica

esquema

- Los ángulos torsionales  $\phi$  (que definen la rotación sobre los enlaces) varían de forma distinta, ya que dicha rotación ha de pasar una barrera periódica. Por tanto, su variación se modela a través de una función periódica del tipo:

$$V_{torsión} = \frac{1}{2} k_n (1 + \cos(n\phi - \gamma)) \quad (25)$$



fórmula

representación gráfica

esquema

Siendo  $k_n$  la barrera de rotación,  $n$  la periodicidad (número de veces que aparece la barrera en un giro completo) y  $\gamma$  la fase (la localización de la primera barrera).

Adicionalmente se pueden modelar los puentes de hidrógeno de forma específica, a través de un potencial análogo al de Lennard-Jones con modificaciones de los parámetros. Esta consideración explícita de los puentes de hidrógeno aparecía en la primera versión del campo de fuerzas de AMBER, en AUTODOCK o en GRID; en otros programas su efecto ya está incluido indirectamente en los términos electrostáticos y de van der Waals. También hay que considerar los llamados ángulos torsionales impropios, que se refieren a la planaridad de un sistema de cuatro átomos en los cuales uno de ellos, en posición central, está enlazado a los otros tres. Estos pueden ser modelados de forma análoga a los ángulos torsionales (a través de potenciales periódicos, como sucede en CHARMM o GROMOS) o bien a través de potenciales harmónicos. Por último, existe la posibilidad de simplificar la representación molecular mediante los modelos *united atom*, que no consideran explícitamente los hidrógenos no polares, de forma que sus características (masa, carga) se añaden al heteroátomo al que estén unidos. Esta aproximación se usa en los programas GROMOS, GRID y AUTODOCK)

El sumatorio de los términos formulados en las ecuaciones (21-25) da lugar a la expresión de la energía potencial de un sistema molecular:

$$V_{pot} = \sum_{enlaces} V_{enlace} + \sum_{ángulos} V_{ángulo} + \sum_{torsiones} V_{torsión} + \sum_{r < r_{cut-off}} (V_{el} + V_{vdW}) \quad (26)$$

El modo de formular cada uno de los términos comentados anteriormente, el método de obtención de cargas parciales y parámetros Lennard-Jones, el número de *atom-types* y el conjunto de parámetros (constantes  $k$  y valores de equilibrio en cada ecuación 21-25) constituye un campo de fuerzas de mecánica molecular (*force-field*). Para una revisión de las distintas formulaciones posibles, véase la referencia [82]. Las ecuaciones (21-25) corresponden al campo de fuerzas parm94,[41] implementado en AMBER[83] y que ha sido el utilizado en los Artículos II y III (en éste último se ha utilizado la versión implementada en el programa de dinámica molecular Q[84]); en este *force-field* se consideran 48 *atom-types* y la obtención de cargas se hace por el procedimiento RESP.[78]

Otros *force-fields* utilizados en esta tesis son el desarrollado por Peter Goodford en el programa GRID[40] (Artículo II) y el implementado en el programa de docking AUTODOCK[85] (Artículos II y III). El objetivo de estos programas es la evaluación de interacciones no enlazantes ligando-receptor ya que, salvo raras excepciones en las que se produce un enlace covalente, son las únicas interacciones que se dan entre un fármaco y su diana terapéutica. La Sección 2.4 describe las características adicionales de estos dos *force-fields* al explicar los potenciales de interacción molecular.

### 2.3. Dinámica molecular

Las aproximaciones de mecánica molecular permiten la evaluación rápida de la energía de una determinada geometría molecular. Pero las moléculas no permanecen fijas en sus geometrías óptimas, sino que pueden fluctuar y moverse a lo largo del tiempo, adaptándose a su entorno y respondiendo a perturbaciones térmicas. Son precisamente estas adaptaciones las que permiten las interacciones fármaco-receptor, y por tanto el estudio dinámico de las moléculas es de vital importancia en los métodos computacionales de diseño de fármacos.

La metodología de dinámica molecular nace de la segunda ley de Newton del movimiento, de forma que la aceleración de una partícula  $i$  puede expresarse en función de su masa  $m$  y la fuerza  $F$  que actúa sobre ella:

$$a_i = \frac{\vec{F}_i}{m_i} \quad (27)$$

La fuerza se puede calcular a partir del gradiente de la energía potencial:

$$\vec{F}_i = -\frac{\partial V_i}{\partial r_i} \quad (28)$$

Si discretizamos el tiempo en fracciones de tiempo (*time step*)  $\delta t$ , y consideramos que los átomos tienen unas velocidades iniciales  $v_0$ , es posible calcular, tras la aplicación de una fuerza, las nuevas posiciones atómicas ( $r_i$ ) y velocidades ( $v_i$ ) para el momento  $(t + \delta t)$ . Para conseguir la trayectoria molecular es preciso utilizar unos métodos de integración que permitan el cálculo sucesivo de dichas posiciones y velocidades. El método de integración más común es el algoritmo de Verlet:<sup>[86]</sup>

Para las posiciones:

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2 \quad (29)$$

Donde la aceleración ( $a$ ) se calcula a partir de la ecuación (27)

Para las velocidades:

$$v(t) = \frac{[r(t + \delta t) - r(t - \delta t)]}{2\delta t} \quad (30)$$

Las velocidades iniciales se asignan aleatoriamente a través de la distribución de probabilidad de Maxwell-Boltzman.

Existen modificaciones de este método, como el algoritmo *leap-frog*, implementado en el programa Q<sup>[84]</sup> utilizado en el Artículo III.

Es importante el control de la temperatura, para asegurarse de cumplir el principio de conservación de la energía. Ello se suele conseguir a través del acoplamiento del sistema a un baño termal, el llamado baño de Berendsen.<sup>[87]</sup> Otro factor limitante es el *time step*, que no puede ser demasiado grande para no aumentar el error del cálculo, que se propaga a lo largo de la simulación. Típicamente es del orden del femtosegundo (fs), lo que hace que las dinámicas moleculares de un sistema biológico, constituido por miles de átomos, no superen el orden de unos pocos nanosegundos (ns). Un truco utilizado para ahorrar tiempo de cálculo es considerar rígidos los enlaces en los que los átomos de hidrógeno están

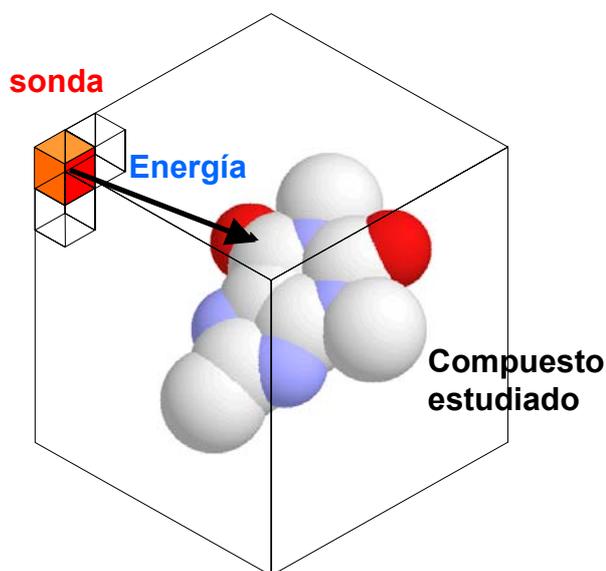
implicados (algoritmo SHAKE), aunque ello implica aumentar el *time step* de forma que sea mayor que la frecuencia de vibración de estos enlaces.<sup>[88]</sup>

Los estudios de dinámica molecular dan cuenta de varias propiedades de un sistema, como son:

- Estabilidad del sistema: desviación con respecto a la geometría inicial, frecuencia de interacciones no enlazantes. Se puede evaluar así la estabilidad de un complejo molecular, como se muestra en el **Artículo II**
- Exploración conformacional: permite al sistema explorar geometrías posibles del sistema de una forma más exhaustiva que mediante los métodos de optimización geométrica basados en el gradiente. Es posible almacenar las distintas conformaciones exploradas para, por ejemplo, realizar experimentos de paralelos de *docking*.<sup>[35]</sup> Por desgracia, los cambios conformacionales implicados en fenómenos biológicos, como la activación de un receptor tras la unión de ligando, ocurren a escalas de tiempo mucho mayores que las que podemos simular. A este objeto se han desarrollado campos de fuerza simplificados,<sup>[89,90]</sup> que al reducir la escala de mínima representación molecular de átomos a residuos, permiten simular periodos de tiempo mayores.
- Cálculo promedio de propiedades moleculares: a partir de un muestreo sistemático de valores de la propiedad que se quiera medir. Esta es la base de los cálculos de energía libre de *binding* por el método LIE,<sup>[47]</sup> (Sección 1.1.4) utilizado en el **Artículo III**.

## 2.4. Potenciales de interacción molecular

Uno de los descriptores químicos más útiles en el diseño de fármacos asistido por ordenador es la energía de interacción de una molécula con un grupo químico determinado, evaluada en puntos del espacio alrededor de la molécula estudiada. Nuestro grupo de investigación tiene una larga tradición en el desarrollo y empleo de este tipo de descripciones moleculares, lo que se refleja en esta tesis en el **Artículo II** y los apéndices II-IV.



**Figura 4:** Esquema de una molécula englobada en una rejilla. Cada nodo de dicha rejilla presentará un valor de energía potencial de interacción.

Como ya he introducido al hablar del 3D-QSAR, para la obtención de esta descripción se discretiza el espacio alrededor de la molécula dividiéndolo en una rejilla (*grid*) y se evalúa la energía de interacción entre la molécula y la sonda química en cada uno de los puntos de la rejilla, tal y como se muestra en la figura 4. La evaluación energética puede hacerse mediante métodos cuánticos, como sucede muy a menudo en el caso del potencial electrostático molecular (MEP) o utilizando las leyes de la mecánica molecular y diversas sondas diferentes al protón, obteniendo entonces potenciales de interacción molecular (MIP). A continuación paso a describir brevemente la obtención de uno y otro potencial:

- El valor del MEP en un punto en el espacio podría ser definido como la energía necesaria para trasladar un protón desde el infinito hasta dicho punto. Se puede calcular de forma rápida y simple mediante la aproximación de cargas puntuales:

$$V(r) = \sum_i^{\forall \text{núcleos}} \frac{\delta_i}{|R - R_i|} \quad (31)$$

Siendo R el punto de la rejilla donde se calcula el MEP ( $V$ ) y  $R_i$  la posición del núcleo de cada átomo  $i$ , que exhibe una carga neta  $\delta_i$

No obstante, es mucho más correcto su cálculo considerando la función de onda, a través de la expresión:

$$V(r) = \sum_i^{\forall \text{núcleos}} \frac{Z_i}{|R - R_i|} - \int \frac{\rho(r)}{|R - r|} dr \quad (32)$$

Donde  $Z_i$  son las cargas nucleares y  $\rho(r)$  es la función de densidad electrónica, la cual se integra en todos los puntos del espacio ( $r$ ).

Los paquetes cuánticos, como GAUSSIAN<sup>[66]</sup> o GAMESS<sup>[67]</sup>, permiten calcular el MEP de una molécula mediante la expresión (32).

- El MIP se obtiene evaluando la interacción de la molécula con diferentes sondas que representan grupos químicos de su contraparte. La forma más extendida del cálculo de esta propiedad es a través del programa de mecánica molecular GRID.<sup>[40]</sup> En este programa, el *force-field* con que se evalúa la interacción intermolecular en cada punto de la rejilla se expresa a través del sumatorio de tres componentes:

$$V_{pot} = V_{el} + V_{vdW} + V_{H-bond} \quad (33)$$

Donde  $V_{el}$  y  $V_{vdW}$  se refieren al valor de los términos electrostáticos y de van der Waals (ver Sección 2.2). El nuevo término  $V_{H-bond}$  modela los puentes de hidrógeno. Debido a que un puente de hidrógeno depende de la distancia del par de átomos electroattractores implicados, pero también del ángulo  $\theta$  formado entre los tres átomos implicados, la expresión toma la forma:

$$V_{H-bond} = \frac{M}{r^m} \frac{N}{r^n} \cos^d \theta \quad (34)$$

La expresión es análoga al potencial de Lennard-Jones pero con exponentes  $m = 6$  y  $n = 4$ , siendo M y N otras constantes determinadas para este potencial y el exponente  $d = 4$ .

Cada sonda, que representa las características de un determinado grupo químico, se considera como un pseudo-átomo al que se le asignan sus correspondientes parámetros en el campo de fuerzas.

De forma análoga, el programa de *docking* AUTODOCK incluye un campo de fuerzas para la evaluación de interacciones no enlazantes (van der Waals, electrostática y puente de hidrógeno), más la contribución a la energía de interacción de la parte entrópica, debida al número de enlaces rotables ( $N_{tor}$ ) del ligando y a un término de solvatación  $V_{sol}$ . Cada uno de estos términos se pondera mediante un coeficiente empírico ( $\Delta G_i$ ), obteniéndose la expresión:

$$\Delta G = \Delta G_{vdW}V_{vdW} + \Delta G_{el}V_{el} + \Delta G_{H-bond}V_{H-bond} + \Delta G_{tor}N_{tor} + \Delta G_{sol}V_{sol} \quad (35)$$

Los parámetros necesarios para calcular los tres primeros términos del sumatorio se derivan del campo de fuerzas de AMBER, utilizándose en este caso potenciales 12-10 para los puentes de hidrógeno.

El empleo de uno u otro potencial dependerá nuevamente del tipo de procesos químicos que se pretenda estudiar. Por ejemplo, una correcta descripción de un estado de transición sólo se obtendrá a través de métodos cuánticos.<sup>[63]</sup>

## 2.5. Métodos estadísticos

En el tratamiento computacional de datos biológicos nos encontramos a menudo con un gran conjunto de “objetos” (ligandos, distintas conformaciones de una molécula) que se caracterizan mediante un conjunto de “variables” (descriptores 3D-QSAR, parámetros geométricos). En estos casos es necesaria la utilización de herramientas estadísticas muy a menudo de naturaleza multivariante. En esta Sección comentaré los fundamentos de los métodos estadísticos utilizados en esta tesis, en los Artículos I y II:

- **Análisis de componentes principales** (PCA, *Principal Component Analysis*)

Es una técnica de reducción de variables, en la que las  $p$  variables originales ( $x$ ) son reemplazadas por un número menor de  $m$  nuevas variables ( $t$ ) llamadas componentes principales, de forma que la pérdida de información sea mínima. Ello se consigue expresando cada nueva variable  $t_i$  como un vector combinación lineal de las  $p$  variables originales,

$$\begin{aligned} t_1 &= c_{11}x_1 + c_{12}x_2 + \dots + c_{1p}x_p \\ t_2 &= c_{21}x_1 + c_{22}x_2 + \dots + c_{2p}x_p \\ &\vdots \\ t_m &= c_{m1}x_1 + c_{m2}x_2 + \dots + c_{mp}x_p \end{aligned} \quad (36)$$

imponiendo la condición de ortogonalidad, es decir, las componentes principales no estarán correlacionadas entre sí

Cada nueva variable  $t_i$  acumula, de forma decreciente (desde 1 hasta  $m$ ), el máximo de información original. En teoría se pueden obtener tantas componentes principales como variables originales, si bien con un número reducido  $m \ll p$  se retiene una cantidad suficiente de la información (varianza) de las variables originales.

Matemáticamente, la matriz original de datos  $X$ , compuesta por  $n$  experimentos (moléculas) y  $p$  variables (descriptores), se descompone en dos matrices: la matriz de los *loadings* ( $P'$ , ecuación 36), que contiene los coeficientes de la combinación lineal, y en la que cada vector (fila)  $t$  es una componente principal, que contiene la información sobre la relación entre las nuevas variables y las originales; y la matriz de los *scores* ( $T$ ), que recoge la proyección de las moléculas en el espacio de los componentes principales, es decir, las coordenadas de cada molécula en el espacio de los  $m$  componentes principales, de forma que recoge las relaciones entre las moléculas. Si se añade la matriz de los residuales ( $E$ ), que explica la varianza no explicada por los componentes principales, obtenemos la fórmula general:

$$X = T \cdot P' + E \quad (37)$$

Al reducir un espacio de cientos o miles de descriptores, que pueden contener información redundante a unas pocas variables, (de 2 a 5, normalmente, y totalmente independientes entre sí) la técnica del PCA permite manejar las descripciones multivariantes de una forma intuitiva, manejable y sobre todo interpretable.<sup>[91]</sup> La distribución que adoptan los objetos (moléculas) en el espacio de descriptores y su agrupamiento en conglomerados (*clusters*), se puede visualizar a través de la representación gráfica de la matriz de *scores*. También se puede determinar cómo contribuyen las variables originales a esta distribución, estudiando la relación entre variables originales y componentes principales a través de la representación gráfica de la matriz de *loadings*. La interpretación de esta relación se puede ver facilitada si se realiza una rotación de los componentes principales,<sup>[92]</sup> utilizando por ejemplo el método Varimax.

Una variante del PCA aplicada a los estudios de regresión es el método PLS (*Partial Least Squares*), en el que a las nuevas variables ortogonales (*latent variables*) definidas por combinación lineal de las originales pretenden obtener la mejor predicción posible de un vector de actividades  $Y$ . Es, como se ha comentado, el método estadístico más comúnmente empleado en los modelos 3D-QSAR.<sup>[93]</sup>

- **Análisis de conglomerados** (*Cluster analysis*)

Es una técnica de análisis de datos que, cuando se aplica sobre una serie heterogénea de objetos, identifica grupos homogéneos definidos según una determinada medida de semejanza.<sup>[94]</sup> Se distinguen dos grandes bloques de técnicas de *clustering*, las jerárquicas y las no jerárquicas.<sup>[95]</sup>

- Técnicas jerárquicas: el análisis es iterativo, de forma que en cada paso un conglomerado se divide en dos (técnicas divisivas) o se produce un agrupamiento de dos conglomerados en uno (técnicas aglomerativas). La forma de representar este análisis es a través de dendrogramas, como el que se muestra en la figura 6 del Artículo II. Estas técnicas se usan también en la construcción de árboles.
- Técnicas no jerárquicas: En estas el número de conglomerados deseado se decide a priori, y la clasificación de los objetos dentro de uno u otro conglomerado se hace de forma iterativa, variando la composición de éstos hasta que se alcanza un determinado criterio de estabilidad en la clasificación. Tal es el caso del algoritmo *k-means*, que se explica con mayor detalle en el Artículo I.

Las técnicas de *clustering* se utilizan en química computacional para reducción de datos, análisis de diversidad o selección de representantes, aplicado al estudio de compuestos o en el análisis conformacional. Existen otras técnicas de selección, que no permiten sin embargo la clasificación en grupos de todo el *set* inicial, como el algoritmo llamado “*most descriptive compound*”,<sup>[96]</sup> utilizado en los Artículos I y II.



### 3. ADENOSINA Y SUS RECEPTORES

El interés farmacológico de esta tesis, en concordancia con el proyecto de investigación en que está enmarcada, es la modulación de unos receptores de membrana activados por adenosina, llamados por tanto receptores de adenosina, que pertenecen al grupo de receptores acoplados a proteína G (GPCRs, *G Protein-Coupled Receptors*).

En este apartado comentaré la caracterización bioquímica y funciones de la adenosina, así como la de sus receptores, con una introducción a la biología estructural relativa a los GPCRs, y finalmente hablaré de los ligandos que activan (agonistas) o bloquean (antagonistas) los receptores de adenosina.

#### 3.1. Adenosina: funciones, biosíntesis y degradación

Químicamente, la adenosina es un nucleósido, compuesto por una base púrica (adenina) y un azúcar (ribosa), unidos a través de un enlace  $\beta$ -*N*-glicosídico. La estructura química y su numeración estándar se muestran en la figura 5.

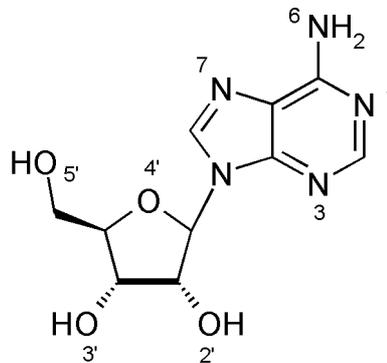


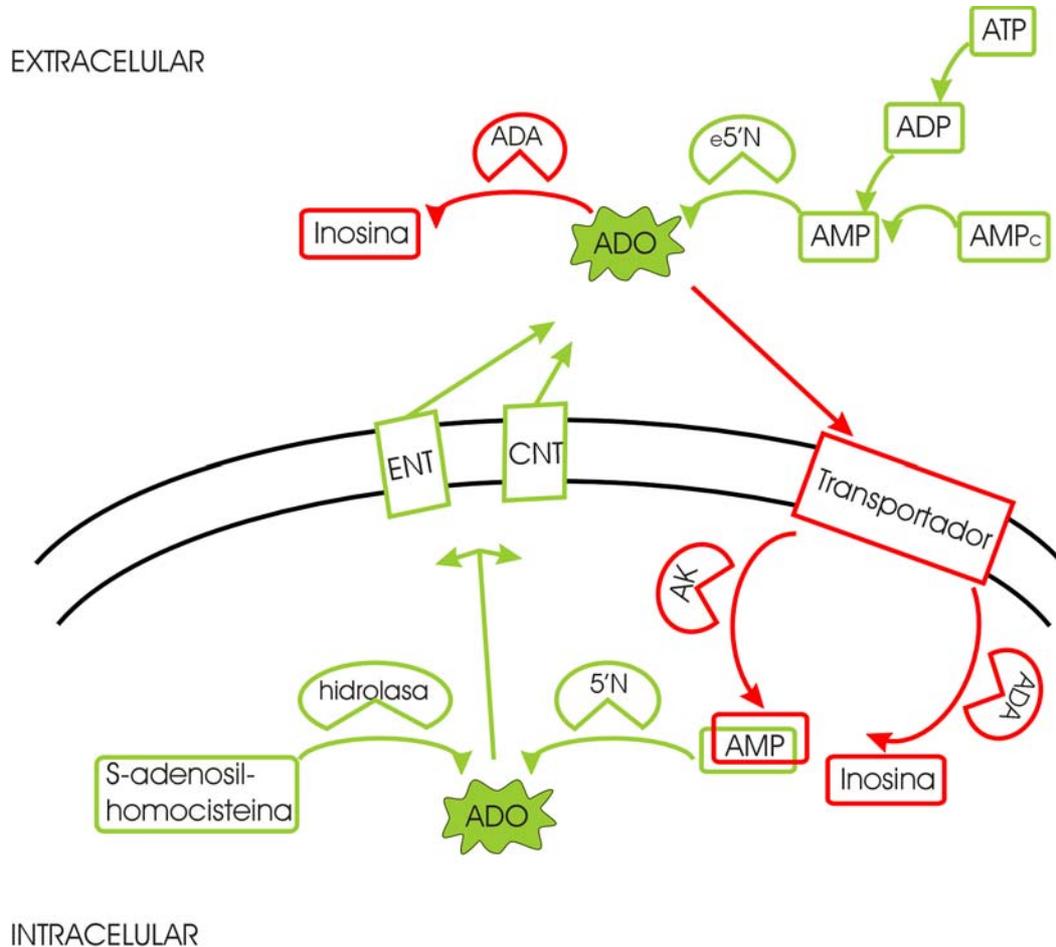
Figura 5: Adenosina

La adenosina está implicada en diversas funciones bioquímicas, ya sea como nucleósido, como nucleótido (es decir, cuando aparece fosforilado en la posición 5') o formando parte de moléculas más complejas. Su función bioquímica se puede dividir en tres grandes grupos:

- Arquitectura de ácidos nucleicos: constituye uno de los cuatro nucleótidos que constituyen los ácidos desoxirribonucleico (ADN) y ribonucleico (ARN), las macromoléculas que almacenan la información genética.
- Función intracelular: diversos derivados de la adenosina tienen una importante función intracelular: el nucleótido ATP (adenosin-trifosfato) es la principal reserva energética de la célula, mientras que los derivados di- (ADP) y monofosforilados (AMP y su derivado ciclado, cAMP) ejercen de segundos mensajeros en diversas rutas bioquímicas de transmisión intracelular de señales. También aparecen nucleótidos de adenosina como cofactores de multitud de enzimas. (NADP, FAD o ATP)
- Función reguladora: la principal función de la adenosina a nivel extracelular es una función reguladora del metabolismo, a través de su interacción con una serie de

*receptores de adenosina*. (ARs, *adenosine receptors*) Esta es la función en la que se centra la presente tesis.

Las rutas de biosíntesis, regulación y degradación de la adenosina se esquematizan en la figura 6.<sup>[97]</sup>



**Figura 6:** Biosíntesis (rutas en verde), degradación (rutas en rojo) de la adenosina. Abreviaturas: ADO (adenosina); ADA (adenosin-deaminasa); AK (adenosin-kinasa); 5'N (5'-nucleotidasa); e5'N (ecto-5'-nucleotidasa).

La biosíntesis puede llevarse a cabo en el interior o en el exterior de la célula. En el primer caso, la adenosina se puede obtener por defosforilación del AMP o bien por hidrólisis de *S*-adenosyl-homocisteína. El transporte hacia el medio extracelular se realiza generalmente por difusión facilitada mediante proteínas ENT, si bien en algunos tejidos existe adicionalmente un transporte en contra de un gradiente de concentración, con la participación de proteínas CNT. La biosíntesis extracelular de adenosina se basa en las rutas bioquímicas de degradación de nucleótidos de adenina. Una vez ejercida su función extracelular de activación de ARs, la adenosina es recaptada a través de transportadores y degradada en el interior de la célula por fosforilación o por desaminación, aunque ésta última también puede tener lugar en el medio extracelular. En la regulación de los niveles extracelulares de adenosina también están implicados otros receptores, como los D<sub>1</sub> de

dopamina o los de NMDA (*N*-metil-D-aspartato), estimulando ambos la liberación de adenosina al exterior de la célula.

La regulación farmacológica de las rutas bioquímicas de la adenosina es motivo de investigación, ya que permite modular los efectos extracelulares de dicha molécula. Así, fármacos como el dipiridamol o el dilazep poseen propiedades vasodilatadoras y de antiagregantes plaquetarios a través del bloqueo de la recaptación de adenosina, con el consiguiente incremento de sus niveles extracelulares. Por otra parte, la inhibición de la degradación de adenosina mediante agentes bloqueantes del enzima adenosina-kinasa (AK) puede dar lugar a nuevos fármacos contra el dolor y la inflamación.<sup>[98]</sup> Esta acción puede ser reforzada por inhibidores de la otra enzima que degrada adenosina, la adenosina deaminasa (AD). En cualquiera de estos casos se produce una regulación indirecta de la acción de la adenosina sobre alguno de sus receptores.<sup>[99]</sup>

### 3.2. Receptores acoplados a proteínas G

De todos los grupos de receptores celulares (proteínas encargadas de la transmisión de señales bioquímicas desde el exterior al interior de la célula), los receptores acoplados a proteína G (GPCRs) constituyen el grupo más numeroso, constituyendo aproximadamente al 1% de las proteínas del genoma humano, lo que, unido con la gran diversidad de funciones que desarrollan, justifica el término de *superfamilia* de receptores. Respecto a su interés en farmacología, una sola cifra parece suficientemente significativa: el 50% de los fármacos que están actualmente en el mercado tienen como diana terapéutica a un GPCR.<sup>[100,101]</sup>

Las características bioquímicas que definen a un GPCR son las siguientes:

- Estructura: son proteínas transmembrana cuya estructura terciaria es de siete  $\alpha$ -hélices unidas por tres *loops* (lazos) extracelulares y tres intracelulares. El fragmento N-terminal de la proteína está situado en la cara extracelular, y posee zonas de glicosidación, mientras que el C-terminal se encuentra en la cara intracelular, con una zona rica en residuos que puedan ser fosforilados.
- Efecto: el receptor se une a un ligando específico, lo cual promueve su activación, necesaria para la transmisión de la señal al interior celular. El primer paso de este proceso es común a los GPCRs y consiste en la disociación de las subunidades de una proteína unida a GTP, la llamada proteína G.

Cada grupo de receptores activado por un mismo grupo químico de ligandos forma una familia, definiéndose una subfamilia para cada grupo de receptores activados por un mismo ligando endógeno. La variabilidad química y estructural de ligandos es inmensa, existiendo GPCRs para proteínas (feromonas), péptidos (glucagón), nucleótidos (ATP), catecolaminas (serotonina), o moléculas tan pequeñas como el calcio,<sup>[102]</sup> incluso la luz puede ser considerada como un ligando, en el caso de las rodopsinas. Esta diversidad, unida a la gran variabilidad de secuencia, hace necesaria una primera clasificación en seis grandes familias o clases de GPCRs. A partir de este punto me centraré en la primera de estas clases (clase A), los receptores tipo rodopsina (*rhodopsin-like*), a la que pertenecen los receptores de adenosina así como la mayoría de los GPCRs de interés terapéutico. La clasificación completa de GPCRs se recoge en la base de datos GPCRDB<sup>[103]</sup> (<http://www.gpcr.org/7tm/index.html>)

EL estudio estructural de estos GPCRs es complicado, y siempre ha venido de la mano de la modelización molecular, tratando de obtener modelos principalmente del sitio de unión que puedan ser de ayuda en el diseño de nuevos ligandos. Los patrones estructurales son escasos, pero han ido mejorando con los años. Así, los primeros modelos partían de la estructura de bacteriorrodopsina,<sup>[104]</sup> un receptor de siete hélices transmembrana pero no acoplado a proteína G. Posteriormente, como se ha comentado en la Sección 1.1.2, se publicó una geometría para la traza de los carbonos alfa de las hélices transmembrana, derivada del alineamiento múltiple de GPCRs conocidos y de la estructura de difracción electrónica de rodopsina.<sup>[16]</sup> Otra aproximación para obtener dicho modelo, con una base más biofísica, se basó en los cálculos de distancias y puentes de hidrógeno entre residuos.<sup>[105]</sup> Actualmente, el mapa de rayos X de la rodopsina bovina<sup>[13]</sup> constituye el patrón estructural para obtener modelos de cualquier GPCR de la clase A.

La otra fuente de información experimental, fundamental para la obtención de modelos tridimensionales de GPCRs, son los estudios bioquímicos disponibles sobre el receptor de interés. Destacan los experimentos de mutagénesis, en los que la estructura del receptor se varía mediante mutaciones, tras lo cual se estudia el cambio de afinidad o actividad registrado para un determinado ligando. La interpretación de estos estudios es compleja, ya que en el caso de que se registre una caída en la actividad será difícil determinar si el cambio en la estructura del receptor estaba directamente relacionado en contactos con el ligando, o bien desestabilizaba la conformación del receptor que debía unir al ligando.<sup>[106]</sup> Se pueden realizar dos tipos de mutaciones:

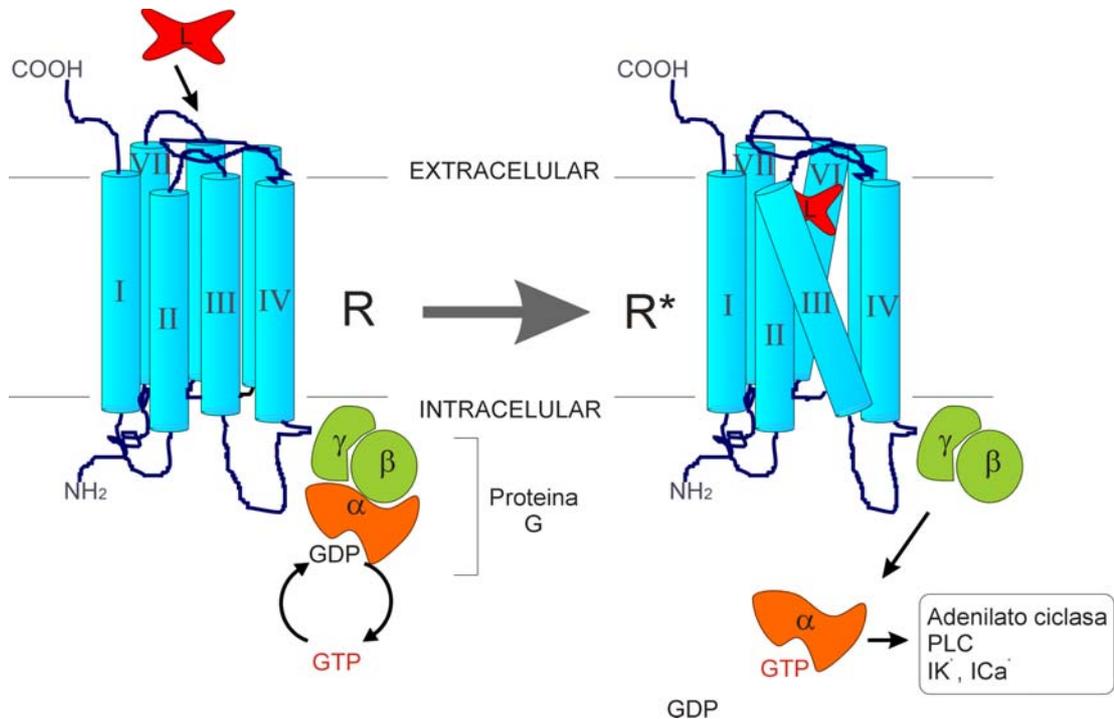
- Construcción de receptores quiméricos: si se sustituye una secuencia de aminoácidos del receptor por la secuencia análoga de un receptor relacionado
- Mutagénesis puntual dirigida (*site-directed*): se muta un único aminoácido del receptor de interés. Los cambios pueden ser conservadores (si el nuevo residuo tiene características similares al original, como la sustitución de leucina por isoleucina) o no conservadores (si las características de ambos residuos son radicalmente distintas, como valina por lisina). La sustitución estándar se hace con alanina, un residuo neutro y pequeño. La interpretación del resultado depende en gran parte del tipo de sustitución que se haya hecho.<sup>[107]</sup>

La importancia de estos estudios es tal que se ha creado una base de datos sobre todos los mutantes publicados de GPCRs y los efectos biológicos de dichas mutaciones denominada GRAP (<http://tinygrap.uit.no/GRAP/>).<sup>[108]</sup> A través de los estudios de mutagénesis se ha hipotetizado sobre los residuos que componen el sitio de unión de un determinado receptor. Tal es el caso de un aspártico en la hélice 3, conservado en los receptores de catecolaminas y sobre el que se puede afirmar con certeza que es el responsable de la interacción por puente salino con el grupo amina de los ligandos.<sup>[109,110]</sup> También se ha revelado el papel estructural de algunos residuos conservados en GPCRs, como la interacción entre las hélices 2 y 7 a través de un par conservado aspártico-asparagina.<sup>[111,112]</sup>

Otras técnicas que han tenido gran importancia en el estudio estructural de GPCRs son el *site-directed spin labeling* (SDSL),<sup>[113,114]</sup> *substituted accessibility cysteine method* (SCAM)<sup>[115]</sup> o la unión covalente con ligandos fotoafines.<sup>[116]</sup>

### 3.2.1. Activación y regulación de un GPCR

La activación y regulación de un GPCR, a pesar de la gran diversidad química de los ligandos de estos receptores, sigue un patrón común. La activación del receptor desencadena una serie de procesos intracelulares que aparecen esquematizados en la figura 7.



**Figura 7:** Esquema del proceso de activación de un GPCR. El ligando (L) estabiliza la conformación activa del receptor (R\*), lo que estimula la dimerización de la proteína G la cual ejercerá distintas funciones intracelulares dependiendo de su naturaleza.

Primeramente el receptor, en estado latente (R), se encuentra unido por su cara intracelular a la proteína G, la cual es un trímero compuesto por las subunidades  $\alpha$ ,  $\beta$  y  $\gamma$ . En este estado, la proteína G une al nucleótido GDP. Al activarse el receptor, se produce un cambio conformacional en éste (R\*) que favorece el intercambio de GDP por GTP en la proteína G, lo cual desemboca en la disociación de sus subunidades  $\alpha$  y  $\beta\gamma$ . Este cambio conformacional no tiene por qué iniciarse de igual manera en cualquier receptor, ya que la naturaleza de los ligandos que favorecen dicho cambio es muy diferente; lo que sí parece común son las etapas finales del cambio conformacional, lo que se ve corroborado por una mayor conservación de secuencia en el lado intracelular de los receptores con respecto al lado extracelular en la clase A de GPCRs.<sup>[117]</sup> Durante la activación parece estar implicado un movimiento de la hélice 6 para alejarse de la hélice 3, así como parece que también juegan un papel importante el motivo Asp/Glu-Arg-Tyr/Trp en la hélice 3 y un aspártico en la hélice 2, todos ellos conservados en la clase A de GPCRs.<sup>[101]</sup> Una vez que la proteína G se ha disociado, se produce la activación de segundos mensajeros intracelulares. Como se muestra en la tabla 2, existen distintas proteínas G dependiendo de la naturaleza de la

subunidad  $\alpha$ , y cada una de ellas promueve una acción intracelular distinta. También la subunidad  $\beta\gamma$  puede promover la activación de alguna ruta bioquímica, como la regulación de canales de  $K^+$ , o de las enzimas adenililciclase y fosfolipasa C.<sup>[118]</sup>

**Tabla 2:** Tipos de proteína G y funciones intracelulares reguladas por cada una de ellas. El símbolo  $\uparrow$  indica activación,  $\downarrow$  indica inhibición y  $\uparrow\downarrow$  indica regulación (adaptado de la referencia [118]).

Subunidad $\alpha$	Adenilato ciclase	Corrientes de iones		$PLC_{\beta}$	Intercambio $Na^+/H^+$	$GMP_c$ - fosfodiesterasa
		$Ca^{2+}$	$K^+$			
$\alpha_s$	$\uparrow$	$\uparrow$				
$\alpha_i$	$\downarrow$		$\uparrow$			
$\alpha_o$		$\downarrow$				
$\alpha_q$				$\uparrow$		
$\alpha_{13}$					$\uparrow$	
$\alpha_t$						$\uparrow$
$\alpha_{olf}$	$\uparrow$					
Subunidad $\beta-\gamma$	$\uparrow\downarrow$		$\uparrow\downarrow$	$\uparrow\downarrow$		

Dependiendo de la línea celular, así como de las condiciones, un mismo receptor puede promover la activación de más de un tipo de proteínas G. Ello se debe a fenómenos de cambios postranscripcionales en la expresión del receptor (*splicing*) en la zona de unión a proteína G y también a la asociación de varias unidades de GPCR para formar oligómeros.<sup>[100]</sup> A pesar de esto, se han realizado intentos de correlacionar la secuencia de un receptor con la proteína G que éste ha de unir, sin tener en cuenta los efectos del *splicing* alternativo o la oligomerización.<sup>[119]</sup>

Una vez activada la ruta intracelular que dará lugar a la respuesta de la célula ante la señal del medio extracelular (que ha venido en forma del *agonista* del receptor), se ha de parar esta respuesta. Es importante resaltar que un solo receptor puede activar consecutivamente a múltiples unidades de proteína G, cada una de las cuales a su vez actúa repetidamente sobre el segundo mensajero, realizándose una *amplificación de la señal*. Así, existen dos formas de regulación: por un lado, la subunidad  $\alpha$  de la proteína G posee actividad GTPasa, con lo que hidroliza GTP a GDP, lo cual la vuelve inactiva y favorece la reunificación de las tres subunidades  $\alpha$ ,  $\beta$  y  $\gamma$ . Por otro lado, después de un tiempo de actividad del receptor, actúa sobre él una quinasa de GPCR (GRK), fosforilando ciertos residuos del fragmento carboxi-terminal lo que conduce a la inactivación del receptor. Este

fenómeno, denominado *desensibilización* del receptor, es responsable de la disminución de eficacia de algunos agonistas de GPCR tras un periodo de tratamiento, como pasa con los agonistas  $\beta$ -adrenérgicos administrados como broncodilatadores.<sup>[118]</sup>

Volviendo al proceso inicial de la activación, una pregunta es aun motivo de debate entre los farmacólogos: ¿es el agonista el que promueve un cambio de conformación en el receptor, o por el contrario este cambio de conformación es espontáneo y lo que hace el agonista es estabilizar la conformación activa? Parece bastante claro que en ausencia de agonista el receptor puede adoptar la conformación activa, lo que se conoce como *activación constitutiva*.<sup>[120]</sup> Ello ha dado lugar a unos modelos de activación del receptor en los que el *agonista* se une a la conformación activa del receptor, estabilizándola, ya que desplaza el equilibrio termodinámico hacia la formación de esta conformación.<sup>[121]</sup> Así aparece el concepto de *agonista inverso*, que define a las moléculas que estabilizan preferentemente la conformación inactiva del receptor, mientras que los *antagonistas* clásicos simplemente bloquean por igual cualquier conformación. No obstante, también se han desarrollado modelos recientes de activación en los que es la unión del agonista la que induce el cambio conformacional.<sup>[122]</sup>

### 3.3. Clasificación de ARs

La primera evidencia del efecto regulador de la adenosina data de 1929, cuando se documentaron sus efectos sobre el sistema cardiovascular.<sup>[123]</sup> Diversos estudios apuntaban a la existencia de unos receptores específicamente activados por adenosina, y a finales de los años setenta se hizo la primera clasificación de receptores purinérgicos,<sup>[124]</sup> en la que aparecían dos grupos de receptores activados por purinas: el primero de ellos ( $P_1$ ) correspondía a receptores activados por adenosina, mientras el segundo ( $P_2$ ) agrupaba receptores activados por ATP u otros nucleótidos. Los consiguientes trabajos de clonación y purificación de receptores permitieron la identificación de cuatro tipos de receptores de adenosina, llamados  $A_1$ ,  $A_{2A}$ ,  $A_{2B}$  y  $A_3$ , que forman una subfamilia de GPCRs.<sup>[97]</sup> Hasta la fecha, se han acumulado 25 secuencias clonadas en 10 especies distintas. La clasificación actual, recogida en la base de datos GPCRDB,<sup>[103]</sup> muestra a los receptores de adenosina como uno de los dos grupos de receptores para nucleótidos, perteneciente a la clase A (*Rhodopsin like*) de los GPCRs; el otro grupo de receptores para nucleótidos es el de los llamados purinoreceptores, activados por nucleótidos. A pesar de este cambio de nomenclatura, no es infrecuente encontrarse en la literatura a los receptores de adenosina como un grupo de receptores purinérgicos.<sup>[125]</sup>

Los cuatro tipos de receptor de adenosina difieren entre sí, además de en su secuencia, en su función fisiológica. En la tabla 3 se muestra la caracterización de cada tipo de AR según distintos parámetros, que explico a continuación:

**Tabla 3:** Clasificación de los receptores de adenosina.

<i>Tipo</i>	<i>Afinidad</i>	<i>Proteína G</i>	<i>Localización</i>
$A_1$	alta	$G_i$	SNC, corazón, hígado, adipocitos
$A_{2A}$	alta	$G_s$	SNC, corazón, pulmón, adipocitos
$A_{2B}$	baja	$G_s$	Cerebro, intestino delgado, vejiga, pulmón
$A_3$	baja	$G_i$	Pulmón, hígado

- Afinidad del receptor por la adenosina: los receptores  $A_1$  y  $A_{2A}$  poseen alta afinidad por la adenosina, activándose con concentraciones de 30–300 nM. Por el contrario, los receptores  $A_{2B}$  y  $A_3$  son llamados de baja afinidad, necesitando típicamente concentraciones de ligando en el rango micromolar para su activación.<sup>[126]</sup> Estos receptores funcionarían, por tanto, en condiciones de estrés celular. Aunque esta clasificación está muy aceptada, experimentos más recientes indican que el receptor  $A_3$  puede ser activado con los niveles de adenosina similares a los necesarios para activar  $A_1$  y  $A_{2A}$ .<sup>[127]</sup>
- Acoplamiento de segundo mensajero: Los receptores de adenosina pueden estar acoplados principalmente a proteína G inhibitoria ( $G_i$ ), caso de los tipos  $A_1$  y  $A_3$ , o estimuladora ( $G_s$ ) como los  $A_{2A}$  y  $A_{2B}$ . Se han descrito posibles acoplamientos a otras proteínas G, aunque no se sabe si generan un efecto fisiológico importante.<sup>[128]</sup> La relación entre la secuencia del receptor de adenosina y la proteína G con quien interactúa ha sido recientemente investigada.<sup>[119]</sup>
- Localización: la tabla recoge los tejidos donde hay una expresión importante de cada uno de los diferentes tipos de receptor de adenosina.

### 3.4. Interés terapéutico

El principal uso terapéutico de la adenosina es en el tratamiento de urgencia de la taquicardia paroxística y diagnosis de arritmias supraventriculares, a través de su estimulación de receptores  $A_1$ .<sup>[118]</sup> No obstante, en los últimos años se ha revelado un gran interés por la modulación de los diferentes tipos de receptores de adenosina como posible tratamiento de distintos cuadros clínicos. A continuación sigue una enumeración de los principales efectos fisiológicos con potencial terapéutico de los receptores de adenosina:

- Sistema nervioso central (SNC)

La cafeína, posiblemente la sustancia psicoestimulante más consumida en el mundo, es un antagonista de los receptores de adenosina<sup>[127]</sup> y constituye un claro ejemplo de las importantes funciones neuroreguladoras en las que están implicados. La amplia distribución de estos receptores en el SNC hace de ellos una potencial diana terapéutica en distintos procesos patológicos.<sup>[129]</sup>

La activación de receptores  $A_1$  en el sistema nervioso central y periférico produce antinocicepción y analgesia.<sup>[130]</sup> No obstante, la activación indirecta de estos receptores por inhibición de la enzima AK parece un mecanismo más efectivo para controlar el dolor y la inflamación en modelos animales.<sup>[98]</sup>

Los receptores de adenosina interactúan con los receptores de dopamina en las neuronas donde éstos se expresan. Concretamente, existe una regulación negativa de los receptores  $A_1$  sobre los de  $D_1$  de dopamina y de los  $A_{2A}$ AR sobre los  $D_2$ .<sup>[131]</sup> Así, se investigan nuevas vías terapéuticas basadas en antagonistas selectivos  $A_{2A}$  para enfermedades motoras como el Parkinson, promoviendo una reactivación de receptores  $D_2$ .<sup>[132]</sup> Por el contrario, los agonistas  $A_1$  y  $A_{2A}$  podrían tener propiedades anticatalépticas,<sup>[130]</sup> interesantes para contrarrestar los efectos secundarios de las terapias con antipsicóticos tipo haloperidol. También existe regulación por parte del receptor  $A_1$  sobre el receptor metabotrópico de glutamato  $1\alpha$  ( $mGlu_{1\alpha}$ ), lo que hace que los agonistas  $A_1$  tengan propiedades neuroprotectoras en condiciones de sobreestimulación de receptores de glutamato, como sucede en el

caso de las enfermedades de Parkinson o de Alzheimer.<sup>[133]</sup> Otros efectos interesantes son el control de la epilepsia y la inducción del sueño provocada por la activación de los receptores A<sub>1</sub>.<sup>[129]</sup>

- Sistema cardiovascular

El receptor A<sub>1</sub> está implicado en el control del ritmo cardíaco, y ésta es la razón del uso de la adenosina como antiarrítmico en taquicardias paroxísticas (Adenocor®). Su baja vida media (debido a la rápida degradación por las enzimas ADA y AK) y sus potenciales efectos secundarios (por su inespecificidad frente a todos los ARs) motivan la búsqueda de agonistas más selectivos. Por otro lado, los antagonistas A<sub>1</sub> pueden ayudar a combatir la fibrilación que aparece tras la fase aguda del infarto de miocardio.<sup>[134]</sup> También pueden ser útiles para combatir la isquemia cardíaca que aparece tras terapias agresivas como el bypass o el transplante de corazón.<sup>[135]</sup> De todas formas, aún hay controversia sobre cuál es el receptor responsable de los efectos cardioprotectores de la adenosina.<sup>[136]</sup> Los efectos de la activación de receptores A<sub>2A</sub> incluyen la inhibición de la agregación plaquetaria y la vasodilatación.<sup>[137,138]</sup>

Otro efecto de la adenosina sobre el sistema cardiovascular viene dado por la amplia distribución de los receptores A<sub>1</sub> y A<sub>2</sub> en la nefrona, mediando así diversos efectos en el sistema renal que pueden ayudar a controlar el flujo sanguíneo.<sup>[130]</sup> Por otra parte, el antagonismo del receptor A<sub>1</sub> en la nefrona es una terapia potencial contra el fallo renal agudo.<sup>[139]</sup>

- Inflamación

Los receptores de adenosina se expresan tanto en la membrana celular de los neutrófilos como en la del endotelio vascular, implicados ambos en los procesos inflamatorios. La activación de estos receptores regula la producción de los mediadores de la inflamación, aumentándolos en el caso del A<sub>1</sub> o inhibiéndolos en el caso de los receptores A<sub>2A</sub> y A<sub>3</sub>.<sup>[130,140]</sup> Sin embargo, parece que el efecto global de el aumento de niveles de adenosina es antiinflamatorio, como sucede tras la administración de metotrexato o sulfasalacina, que ejercen su acción antiinflamatoria aumentando los niveles de adenosina.<sup>[141]</sup> Por tanto, el tratamiento de desórdenes inflamatorios crónicos como la artritis puede pasar por la activación de receptores A<sub>2A</sub><sup>[142]</sup> y A<sub>3</sub><sup>[140]</sup> o la inhibición del enzima AK.<sup>[143]</sup>

- Inflamación en vías aéreas

Los procesos inflamatorios alérgicos, como el asma, están mediados por la degranulación de los mastocitos en las vías aéreas, lo que produce aumento de mediadores inflamatorios, como la histamina y prostanoídes, y consiguiente broncoconstricción. En estas condiciones los niveles de adenosina aumentan y los receptores A<sub>2B</sub> y A<sub>3</sub> se sobreexpresan en mastocitos y pulmones.<sup>[144,145]</sup> La importancia de este fenómeno queda reflejada en el uso de un antagonista de adenosina, la teofilina, en el tratamiento clásico del asma.<sup>[118]</sup> No obstante, la teofilina es un antagonista no selectivo y la búsqueda de moléculas selectivas A<sub>2B</sub><sup>[144,146]</sup> o A<sub>3</sub><sup>[145]</sup> es una vía prometedora para mejorar los tratamientos actuales de esta enfermedad.<sup>[147]</sup> El presente proyecto se originó en el marco de una línea de investigación multidisciplinar sobre el diseño y la evaluación de nuevos fármacos antiasmáticos activos por vía oral, en la que participaban grupos de la universidad y de la industria.<sup>[148]</sup>

### 3.5. Caracterización estructural

La estructura de los receptores de adenosina es común a la de cualquier GPCR y, como tal, aún no ha sido revelada por técnicas experimentales. Los métodos computacionales de modelado, junto con la interpretación de los datos bioquímicos disponibles, constituyen de momento la única forma de caracterizar estructuralmente estos receptores.

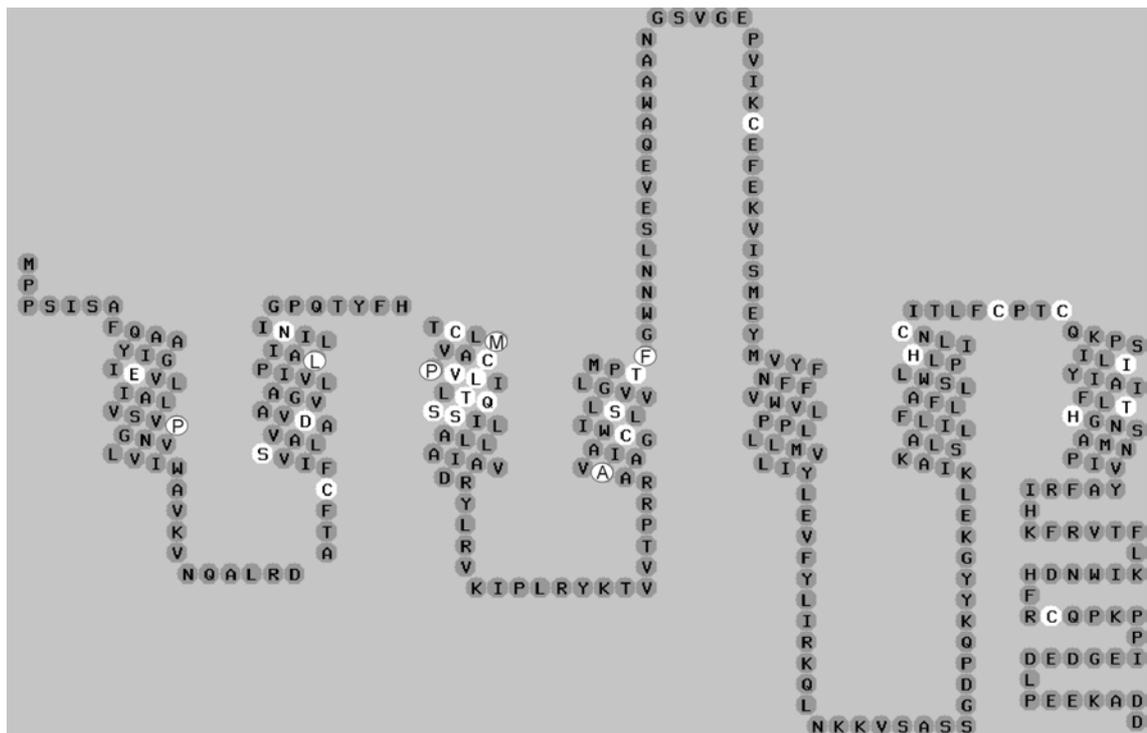
Las técnicas bioquímicas comentadas anteriormente, principalmente los ensayos de mutagénesis dirigida, han revelado una serie de posiciones que son clave bien para la conservación de la estructura del receptor o bien para la unión de ligandos. Los receptores mejor caracterizados mediante estas técnicas son el A<sub>1</sub> y el A<sub>2A</sub>. En la figura 8 se muestra una representación bidimensional del receptor A<sub>1</sub>, apareciendo remarcadas las posiciones que han sido ensayadas por mutagénesis dirigida. Como se ha comentado, la interpretación de estos datos es compleja, y pasa por la construcción de modelos tridimensionales. Existen revisiones sobre la interpretación de todo este conglomerado de datos y modelos,<sup>[97,149]</sup> y en la tabla 4 se muestran los modelos publicados hasta la fecha para cada uno de los receptores de adenosina.

**Tabla 4:** Modelos moleculares de los receptores de adenosina, ordenados según fecha de publicación.

<i>Receptor</i>	<i>Patrón (código PDB)</i>	<i>Método</i>	<i>Referencia</i>
A <sub>1</sub>	bacteriorodopsina <sup>c</sup>	homología	Ijzerman, 1992 <sup>[150]</sup>
A <sub>1</sub>	rodopsina (1BAC)	topología	Bianucci, 1998 <sup>[151]</sup>
A <sub>1</sub>	modelo A <sub>2A</sub> (1MMH)	homología	Rivkees, 1999 <sup>[152]</sup>
A <sub>1</sub>	rodopsina (Baldwin <sup>c</sup> )	topología	Da Settimo, 1999 <sup>[153]</sup>
A <sub>1</sub>	rodopsina b. (1F88)	pseudoreceptor	Bondavalli, 2002 <sup>[154]</sup>
A <sub>2A</sub>	bacteriorodopsina <sup>c</sup>	homología	Ijzerman, 1994 <sup>[155]</sup>
A <sub>2A</sub>	rodopsina (1993 <sup>c</sup> )	topología	Kim, 1995 <sup>[156]</sup>
A <sub>2A</sub>	rodopsina b. (1F88)	topología <sup>a</sup>	Kim, 2003 <sup>[157]</sup>
A <sub>2B</sub>	rodopsina b. (1F88)	topología	Pastorin, 2003 <sup>[158]</sup>
A <sub>3</sub>	rodopsina (Baldwin <sup>c</sup> )	topología	Li, 1999 <sup>[159]</sup>
A <sub>3</sub>	rodopsina b. (1F88)	topología	Baraldi, 2001 <sup>[160]</sup>
A <sub>3</sub>	rodopsina b. (1F88)	topología <sup>a,b</sup>	Jacobson, 2001 <sup>[161]</sup>
A <sub>3</sub>	rodopsina b. (1F88)	homología	Cristally, 2003 <sup>[162]</sup>

<sup>a</sup> Uno o varios *loops* extracelulares fueron modelados. <sup>b</sup> Este trabajo incluye también modelos de receptores mutantes. <sup>c</sup> Estas estructuras, debido a su baja resolución, no fueron publicadas en el *Protein Data Bank*.

**Figura 8:** Diagrama bidimensional del receptor de adenosina A<sub>1</sub>, utilizando una secuencia consenso sobre las distintas especies para las que ha sido clonado. Los residuos que han sido estudiados por mutagénesis aparecen señalados. En la tabla adjunta se detalla la numeración estándar de los residuos estudiados. Se ha usado el siguiente código de colores: rojo: afecta la afinidad de algún ligando en más de 100 veces; amarillo, cambios de 2 a 100 veces; negro, no hay cambios aparentes en la afinidad de ligandos.



TMH I	TMH II	TMH III	TMH IV	TMH V (EL2)	TMH VI (EL3)	TMH VII (NH <sub>2</sub> TER)
<b>Glu1.39</b> <sup>[163]</sup>	Cys2.41 <sup>[164]</sup>	<b>Cys3.25</b> <sup>[164]</sup>	Ala4.43 <sup>[152]</sup>	Cys5.27 <sup>[164]</sup>	<b>His6.52</b> <sup>[165]</sup>	<b>Ile7.39</b> <sup>[166]</sup>
<b>Pro1.48</b> <sup>[152]</sup>	Ser2.45 <sup>[163]</sup>	Met3.27 <sup>[152]</sup>	Cys4.49 <sup>[164]</sup>		Cys6.56 <sup>[164]</sup>	<b>Thr7.42</b> <sup>[167]</sup>
	Asp2.50 <sup>[163]</sup>	<b>Cys3.30</b> <sup>[164]</sup>	Ser4.53 <sup>[163]</sup>		Cys6.61 <sup>[164]</sup>	<b>His7.43</b> <sup>[165]</sup>
	Leu2.60 <sup>[152]</sup>	<b>Pro3.31</b> <sup>[152]</sup>	Thr4.59 <sup>[163]</sup>		Cys6.64 <sup>[164]</sup>	Cys7.74 <sup>[164]</sup>
	Asn2.65 <sup>[152]</sup>	<b>Val3.32</b> <sup>[152]</sup>	Phe4.62 <sup>[152]</sup>			
		<b>Leu3.33</b> <sup>[152]</sup>				
		<b>Thr3.36</b> <sup>[152]</sup>				
		<b>Gln3.37</b> <sup>[152]</sup>				
		Ser3.38 <sup>[163]</sup>				
		<b>Ser3.39</b> <sup>[163]</sup>				

Como se puede ver, las principales diferencias entre estos modelos vienen por la estructura patrón que se ha tomado para modelar el receptor, así como por la técnica empleada para el modelado. Estos modelos han servido para interpretar el sitio y modo de unión de agonistas y antagonistas. La caracterización de este sitio no está completada a partir de los datos experimentales disponibles, si bien parece claro que está localizado en la zona transmembrana, con un eje principal entre las hélices 3, 6 y 7.<sup>[97,149]</sup>

En los Artículos II y III se discute ampliamente sobre los datos disponibles para el receptor A<sub>1</sub> sobre la base de un modelo computacional desarrollado en nuestro laboratorio para éste receptor. Asimismo, en la Sección de **Discusión** se explica el procedimiento de evaluación de datos de mutagénesis seguido en dichos trabajos.

### **3.6. Agonistas y antagonistas**

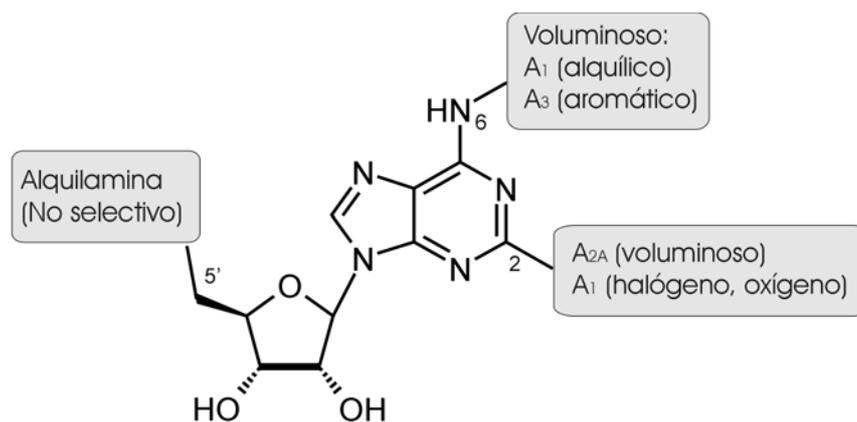
Debido a este amplio potencial terapéutico, pero también a la necesidad de una mejor caracterización farmacológica de los receptores de adenosina, es interesante encontrar ligandos con alta afinidad y selectividad para cada uno de los cuatro tipos de AR.

En esta Sección se mostrará una visión general de las características químicas de los agonistas y antagonistas de receptores de adenosina. Existen revisiones más exhaustivas en la literatura sobre ligandos de receptores de adenosina,<sup>[168]</sup> y en particular sobre sus agonistas,<sup>[99]</sup> sobre sus antagonistas<sup>[126]</sup>, y también sobre los requerimientos estructurales de afinidad/selectividad de cada uno de los cuatro subtipos de ARs.<sup>[145,169,170]</sup>

#### **3.6.1. Agonistas**

Los agonistas de los receptores de adenosina se limitan a modificaciones químicas de la estructura de la propia adenosina. Con estas modificaciones se busca una triple acción: aumento de afinidad, selectividad frente a un receptor concreto y conseguir ligandos estables frente a las enzimas degradativas (ADA y AK). Hasta la fecha han aparecido agonistas selectivos para los receptores A<sub>1</sub> y A<sub>3</sub>, mientras los receptores A<sub>2</sub> aún no tienen ligandos selectivos.<sup>[99,168]</sup>

En la figura 9 se muestran las posiciones del esqueleto de adenosina donde se pueden introducir modificaciones químicas:



**Figura 9:** Modificaciones químicas sobre el esqueleto de adenosina que conducen la obtención de agonistas de los receptores de adenosina.

- Posición N6:** En dicha posición se puede introducir un sustituyente voluminoso, el cual aumenta la afinidad y selectividad por los receptores A<sub>1</sub> y A<sub>3</sub>, mientras que la disustitución disminuye drásticamente la afinidad por cualquier AR. Si el sustituyente es un grupo alquílico, aumenta la selectividad por el receptor A<sub>1</sub>, mientras que los sustituyentes aromáticos son preferidos por el receptor A<sub>3</sub>.<sup>[145]</sup> No obstante, estas selectividades son moderadas y raramente se alcanza una potencia 100 veces mayor por alguno de los dos receptores con dichas modificaciones. Para aumentar la selectividad A<sub>3</sub> es necesario combinar sustituciones en N6 y en 5', como sucede en el ligando AB-MECA (N6-(4-aminobencil)-5'-N-metilcarboxamidoadenosina). Todos los sustituyentes en N6 que resultan efectivos incluyen un ciclo, el cual puede estar directamente enlazado al átomo de nitrógeno N6 (la N6-ciclohexiladenosina o CHA es el ligando más selectivo A<sub>1</sub>), o unido a través de una cadena alquílica. Un ejemplo de este último tipo es la N6-fenilisopropiladenosina (PIA) que posee una marcada estereoselectividad a favor del estereoisómero R.
- Posición 2:** en este caso, sustituyentes voluminosos aumentan la selectividad por el receptor A<sub>2A</sub>. Estos sustituyentes son normalmente heterociclos unidos a la posición 2 por una cadena alquílica, la cual puede contener un fragmento polar como sucede en el caso del ligando CGS21680 (2-[4-[[2-carboxietil]fenil]etilamino]-5'-N-etilcarboxamidoadenosina). Estas sustituciones suelen estar combinadas con sustituciones en la posición 5'. Aun así, la selectividad por el receptor A<sub>3</sub> es baja. Conviene destacar que en éste último receptor estas modificaciones en posición 2 del núcleo de adenosina y las comentadas anteriormente para la posición N6, si bien por separado aumentan la afinidad, no pueden coexistir en una misma molécula.<sup>[99]</sup>

Por último, en el receptor A<sub>1</sub> se permiten sustituciones pequeñas (halógenos, oxígeno), que combinan bien con la sustitución voluminosa en N6 (2-cloro-N6-ciclopentiladenosina, CCPA)<sup>[171]</sup>
- Posición 5':** el grupo hidroxilo en esta posición puede ser reemplazado por una alquilamina, lo que conduce a un aumento general de la afinidad en los cuatro receptores, como sucede en el ligando inespecífico NECA (N-

etilcarboxamidoadenosina).<sup>[172]</sup> El espacio en los receptores para acomodar dicha cadena alquílica es limitado, ya que sólo se toleran sustituciones pequeñas (por lo general, uno o dos carbonos). También se tolera un halógeno en esta posición,<sup>[145]</sup> mientras que 5'-metilseleno o 5'-metiltio derivados generan agonistas parciales para el receptor A<sub>1</sub>.<sup>[173]</sup>

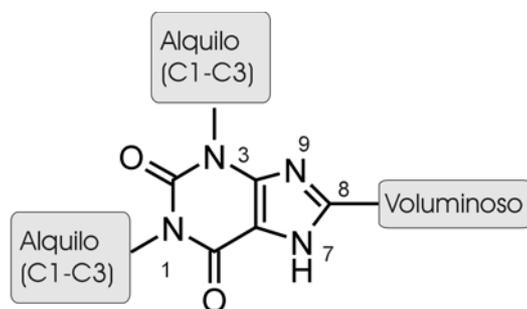
Otras modificaciones posibles incluyen la síntesis de 1-dezaadenosinas,<sup>[174]</sup> siendo éste el único nitrógeno del heterociclo de adenina prescindible para la unión a receptores de adenosina, a pesar de que la actividad se puede ver afectada.<sup>[175]</sup>

Los requerimientos del fragmento ribosa son muy específicos: estereoquímicamente es una β-D-ribofuranosa (ver fig. 9), con conformación bioactiva *anti*, es decir, con el anillo de adenina y la ribosa apuntando en direcciones opuestas, frente a la conformación alternativa (*sin*).<sup>[99]</sup> Las modificaciones en los grupos 2' y/o 3' del anillo de ribosa conducen a un perfil de agonista parcial,<sup>[175,176]</sup> mientras que la eliminación de los dos grupos a la vez convierte a la molécula en antagonista.<sup>[177]</sup> En todos los casos se produce también una disminución de la actividad.<sup>[99]</sup>

Alternativamente, se han sintetizado agonistas que combinan esqueleto de xantina con una ribosa en posición 7, los cuales poseen una moderada afinidad por el receptor A<sub>3</sub>, así como metanocarbo derivados de adenosina conformacionalmente restringidos, que han ayudado a determinar la conformación biológicamente activa del anillo de ribosa.<sup>[178]</sup>

### 3.6.2. Antagonistas

Las moléculas que bloquean la actividad de los receptores de adenosina pertenecen a familias químicas diversas. Tradicionalmente se dividen en dos grandes bloques, la familia de las xantinas y los antagonistas no xantínicos.<sup>[126,168]</sup>



**Figura 10:** Xantina: esquema 2D y numeración, resaltando las posiciones que admiten sustitución.

- Xantinas

Como se ha dicho, la cafeína (químicamente 1,3,8-trimetilxantina) es un antagonista clásico de los receptores de adenosina.<sup>[127]</sup> Se han realizado múltiples modificaciones químicas sobre el esqueleto de xantina con el objetivo de conseguir moléculas selectivas A<sub>1</sub>, A<sub>2A</sub> ó A<sub>2B</sub>. El receptor A<sub>3</sub>, por el contrario, no muestra alta afinidad por los derivados xantínicos, habiendo diferencias interespecie.<sup>[145]</sup> Las modificaciones que modulan la selectividad y afinidad se reducen a 3 posiciones (ver fig 10): en las posiciones 1 y 3 se requiere una cadena alquílica, compuesta

generalmente por entre uno y tres átomos de carbono. En la posición 8, sustituyentes voluminosos aumentan la afinidad por todos los receptores, aunque los requerimientos estructurales dependen del tipo de AR. Así, se ha conseguido una alta selectividad  $A_1$  mediante introducción de un grupo cicloalquilo (DPCPX, 1,3-dipropil-8-ciclopentilxantina). Un grupo muy interesante lo constituyen los congéneros xantina-amina (XAC, *xantine amine congener*) que han dado lugar a múltiples derivados.<sup>[179]</sup>

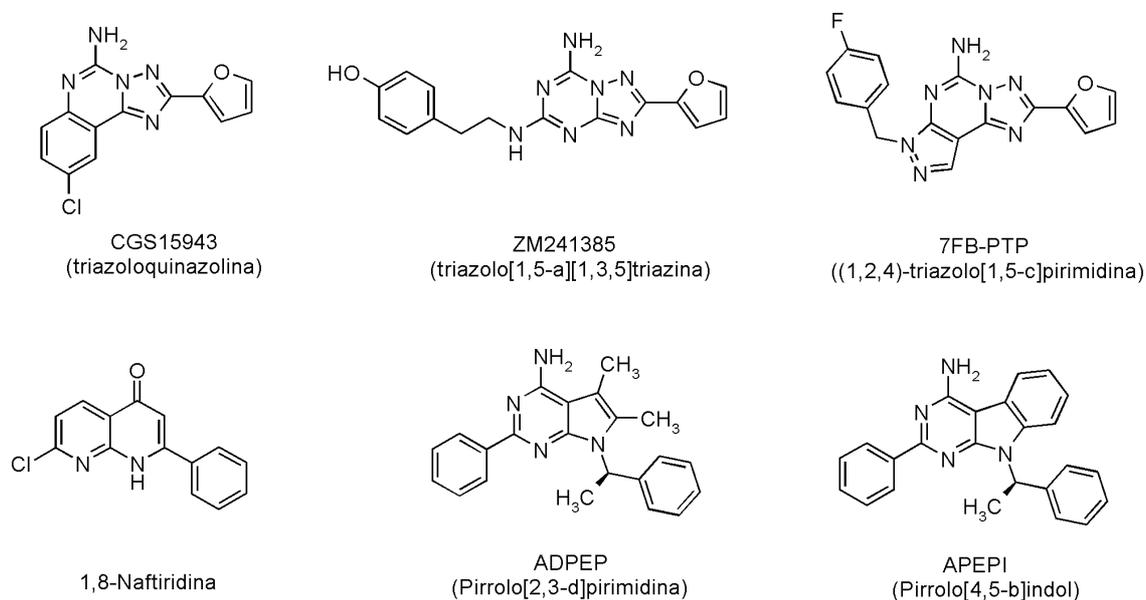
Es interesante la analogía entre el sustituyente C8 de las xantinas y el sustituyente N6 de los derivados de adenosina, lo que ha llevado a la proposición del modelo de superposición C8/N6 entre xantina y adenina.<sup>[180,181]</sup> Este modelo ha sido obtenido también en nuestro grupo de investigación mediante superposición de campos de interacción molecular con el programa MIPSim.<sup>[182]</sup>

Por último, de los tres posibles puntos de aceptación de puente de hidrógeno, al menos dos de ellos son necesarios para una alta afinidad,<sup>[126]</sup> lo que condujo a la síntesis de 9-deazaxantinas como antagonistas de receptores de adenosina.<sup>[183]</sup>

- Derivados no xantínicos

Se ha sintetizado una gran variedad de antagonistas con estructuras heterocíclicas diferentes a la de la xantina, y tan sólo comentaré aquí aquellas familias de compuestos que se han manejado en el proyecto de investigación en el que se enmarca la presente tesis. Las correspondientes estructuras químicas aparecen en la figura 11.

**Figura 11:** Ejemplos de cada familia de antagonistas citada en el texto.



Históricamente, el primer grupo lo constituyen las triazoloquinazolinas, cuyo ejemplo más representativo es el antagonista no selectivo CGS15943.<sup>[184]</sup> Partiendo de este ejemplo se obtuvo la triazolo[1,5-a][1,3,5]triazina ZM241385, antagonista selectivo para los receptores  $A_{2A}$ ,<sup>[185]</sup> y la familia de las triazolopirimidinas.<sup>[186]</sup>

Dentro de los antagonistas selectivos A<sub>1</sub> destaca la familia de las pirrolopirimidinas y pirimidindoindoles<sup>[187]</sup> y los derivados de naftiridina.<sup>[188]</sup> Esta diversidad química de antagonistas hace necesarios estudios teóricos en los que se comparen las distintas familias, basándose en las propiedades químicas responsables de la interacción con el receptor. Existen algunos ejemplos en la literatura,<sup>[187,189]</sup> y en nuestro grupo de investigación estamos llevando a cabo un estudio de estas familias basándonos en comparaciones de campos de interacción molecular con el programa MIPSim. El resultado preliminar de estos estudios se muestra en el **apéndice III**

Otras familias químicas, como las piridinas, flavonoides o isoquinolonas, incluyen antagonistas de los receptores A<sub>3</sub>, receptores cuyo perfil de respuesta frente a antagonistas difiere de los otros ARs.<sup>[160]</sup>

## 4. DISCUSIÓN

La presente tesis se enmarca dentro de una línea de investigación multidisciplinar sobre el diseño, síntesis y evaluación farmacológica de nuevos fármacos con mayor eficacia y menos reacciones adversas en el tratamiento de la rinitis y el asma alérgico, promovida por Almirall-Prodesfarma y en el que, por la parte académica, participaron los departamentos de Química Orgánica y Farmacología de la Universidad de Santiago de Compostela, el departamento de Química Farmacéutica de la Universidad de Bari (Italia) y el Grupo de Investigación en Informática Biomédica del IMIM y de la Universidad Pompeu Fabra. Dicha línea de investigación se enfocó hacia el descubrimiento de antagonistas potentes y selectivos para el receptor  $A_{2B}$  y los correspondientes resultados se han patentado<sup>[148]</sup> o se encuentran en proceso de publicación.

En el marco de dicha línea de investigación se han explorado diferentes estrategias en el ámbito del diseño de fármacos asistido por ordenador. En este contexto, esta tesis se ha ocupado de la caracterización estructural de los receptores de adenosina, sus ligandos y de los complejos ligando-receptor, tratando de aportar información que ayude a entender las relaciones estructura-actividad. Aparte de la utilidad de la presente tesis para la línea de investigación aplicada anteriormente descrita, creemos que sus contribuciones aportan un avance significativo en ámbitos metodológicos como:

- Metodología de modelado de GPCRs,
- Estrategias de simulación y análisis del acoplamiento ligando-receptor
- Métodos para el abordaje de estudios de diversidad molecular.

Siguiendo la clasificación propuesta en la Sección 1 de la Introducción, paso a discutir los aspectos más relevantes del trabajo aquí presentado. Asimismo, se plantean las cuestiones que quedan abiertas en el ámbito abordado por la presente tesis.

### 4.1. Métodos directos (diseño de fármacos basado en receptor):

El receptor  $A_{2B}$  es un receptor de baja afinidad, no existiendo hasta la fecha ligandos específicos y selectivos para él. Este hecho hace que sea el peor caracterizado bioquímicamente, con lo que la construcción directa de un modelo de este tipo de receptor de adenosina no es aconsejable. En nuestro laboratorio hemos optado por estudiar un receptor relacionado, el tipo  $A_1$  de los receptores de adenosina, un receptor de alta afinidad del que se dispone de multitud de datos bioquímicos,<sup>[97,169]</sup> para caracterizar sobre el modelo molecular derivado el sitio y modo de unión de ligandos. Partiendo entonces de esta información, estimamos mucho más fiable la potencial construcción de un modelo del tipo  $A_{2B}$ , en el que se puedan incorporar restricciones derivadas del estudio del receptor  $A_1$ .

La construcción del modelo molecular del receptor humano  $A_1$  de adenosina ha tenido en cuenta una serie de puntos que merecen especial consideración:

- En primer lugar, el tratamiento de datos experimentales, concretamente los datos de experimentos de mutagénesis dirigida, requiere una consideración sistemática de toda la información disponible. Uno de los puntos críticos de dichos experimentos es la interpretación de sus resultados, ya que el descenso de actividad y/o afinidad en un determinado mutante puede ser debido a diversas razones, y no necesariamente al contacto directo de la posición mutada con el ligando.<sup>[106]</sup> A menudo existen diferentes ensayos sobre un determinado mutante, en diferentes

condiciones o con distintos ligandos, y la interpretación del dato resultante varía de un autor a otro. Para intentar minimizar la subjetividad en la interpretación de estos datos, y también al objeto de obtener una visión global de los datos disponibles, en el desarrollo de esta tesis he construido dos bases de datos de experimentos de mutagénesis dirigida con los datos disponibles para los receptores  $A_1$  y  $A_{2A}$ , respectivamente. Estas bases de datos, de las cuales se muestra una entrada en la figura 12, han sido construidas con el programa IsisBase,<sup>[190]</sup> e incluyen los siguientes campos:

- **Ligando:** Se especifica el tipo (agonista o antagonista), su nombre (según las siglas convencionales) y su estructura 2D. Este último campo es especialmente interesante ya que permite la búsqueda por subestructuras y así determinar, por ejemplo, si un determinado esqueleto químico se ve siempre influenciado de la misma forma ante una cierta mutación en el receptor.
- **Receptor:** Se especifica la especie, la posición mutada (según la nomenclatura de GPCRs propuesta en la referencia <sup>[31]</sup>), el aminoácido original y el mutante (en código de tres letras) y el efecto (número de veces que desciende la actividad como consecuencia de la mutación, o el inverso del número de veces que asciende ésta, en caso de incremento de actividad).
- **Referencia:** se incluye la referencia original en formato EndNote.

type	antagonist	specie	bovin
lig_name	XAC	position	6.52
ligan		aminoacid	His
		mutant	Leu
		effect	4.0000
reference	[Olah, 1992 #11]		

**Figura 12:** Entrada de la base de datos de mutantes

De tal forma ha sido posible manejar de forma cómoda y objetiva todo el volumen de datos referido a experimentos de mutagénesis durante la elaboración de esta tesis. Se ha podido comparar el efecto de mutaciones en dos tipos de receptor, a fin de determinar si la función de una posición se conserva o bien es específica de un único tipo. También se ha comprobado la consistencia de los datos o su inconsistencia, en el caso de existir más de un estudio sobre una determinada posición. En cuanto a la naturaleza del ligando, es posible detectar si una mutación afecta específicamente a una familia de compuestos mientras otra familia no ve alterada su farmacología, lo que puede ser una información útil para evaluar modos de unión alternativos.

- La evaluación dinámica de la naturaleza del receptor: las técnicas de dinámica molecular son frecuentemente empleadas al final de la construcción de modelos moleculares a fin de evaluar su estabilidad o la frecuencia de determinadas interacciones intra o intermoleculares. En nuestro caso, además de considerar estos aspectos, el estudio de dinámica molecular ha incluido el análisis de otros factores. Así, hemos determinado el estado de protonación más probable de dos residuos histidina, importantes en la unión de ligandos. Por otro lado, las trayectorias de dinámica molecular se han analizado mediante análisis de componentes principales y posterior selección de representantes, de forma que hemos podido hacer un estudio de las conformaciones visitadas por el receptor considerando únicamente 11 conformeros representativos que cubren adecuadamente el espacio conformacional de interés. En este sentido la metodología empleada y la idea que subyace a ésta son análogas a la empleada en el **Artículo I** para la selección de representantes de una base de datos de compuestos químicos (ver más adelante).
- Localización del sitio de unión: debido a la falta de información sobre el sitio de unión de agonistas a los receptores de adenosina, hemos propuesto una estrategia original basada en un mapeo de posibles sitios de unión de ribosa sobre el receptor estudiado y su comparación con sitios de unión de ribosa recogidos en el PDB. La caracterización de los sitios de unión de ribosa ha sido llevada a cabo mediante técnicas propias de la obtención de farmacóforos, en este caso aplicadas en sentido inverso para detectar la relación espacial entre puntos de la proteína que unen al fármaco. Los resultados han permitido detectar un perfil común en los sitios de unión de ribosa tal como se detalla en el **Artículo II**.

Una vez construido el modelo del receptor  $A_1$  y propuesto su *sitio* de unión de agonistas, el estudio realizado en el **Artículo III** ha tenido como finalidad la determinación del *modo* de unión más probable para la familia de derivados de adenosina. Como se ha explicado en la **Sección 3.6.1** de la introducción, todos los agonistas de receptor de adenosina, salvo raras excepciones, son derivados químicos del ligando natural. Este hecho hace pensar que el modo de unión a receptor ha de ser conservado entre diferentes adenosin-derivados. La estrategia utilizada en este estudio ha tenido en cuenta la búsqueda automática de modos de unión con cuatro agonistas (con el programa de *docking* AUTODOCK), a fin de encontrar una posición de unión común a todos ellos, y en la que además se pudiese explicar la función de los diferentes sustituyentes admitidos por el esqueleto de adenosina. Más aún, en una segunda fase del estudio se han utilizado técnicas avanzadas para la determinación de la energía de unión de ligandos, a fin de evaluar si alguna de las posiciones propuestas respetaba las diferencias experimentales de afinidad existentes entre los ligandos estudiados. La implementación de estas técnicas en nuestro modelo ha sido motivo de una estancia

predoctoral en la Universidad de Uppsala, colaborando con el grupo del Profesor Åqvist, inventor de la técnica LIE empleada en este estudio (ver Sección 1.1.4 de la Introducción). Los resultados muestran que la estrategia seguida en este estudio ha dado como resultado la propuesta de un modo de unión conservado entre todas las moléculas estudiadas, capaz de explicar diferencias de afinidad entre ellas, relacionándolas con diferencias en la estructura química. Este modo de unión puede servir, por tanto, como punto de partida para el diseño de moléculas agonistas más activas, pero también para la determinación del modo de unión de antagonistas según alguno de los modelos de superposición agonista-antagonista propuestos en la literatura.

El objetivo final de esta parte de **métodos directos**, dentro del proyecto de colaboración de diseño de nuevos fármacos, sería la obtención de modelos de todos los tipos de receptor de adenosina, a fin de poder proponer condicionantes estructurales para la selectividad. En este sentido, el trabajo derivado de la presente tesis sobre el receptor A<sub>1</sub> serviría como punto de inicio para la construcción de los otros receptores: por una parte, se ha puesto a punto una metodología de construcción de GPCRs aplicada a un receptor de adenosina basándonos en experiencias previas de nuestro grupo de investigación,<sup>[32,191]</sup> pero incorporando nuevas técnicas en la caracterización de la dinámica y el sitio de unión del receptor (**Artículo II**). Por otro lado, la información derivada del estudio del receptor A<sub>1</sub> puede ser incorporada a la hora de construir receptores homólogos. En cuanto a la caracterización de la selectividad, la metodología empleada en el **Artículo III** para distinguir entre dos posibles formas de unión de agonistas al receptor A<sub>1</sub> puede ser utilizada para evaluar la diferente afinidad de un ligando por dos receptores diferentes, una vez construidos los modelos de éstos.

#### **4.2. Métodos indirectos (diseño de fármacos basado en ligando):**

En el marco de estudios orientados al descubrimiento de nuevos ligandos de los receptores de adenosina se realizó el estudio descrito en el **Artículo I**, cuyo objetivo consistía en diseñar un experimento de química combinatoria mediante el que realizar una exploración sistemática de diversidad molecular en una determinada posición en una molécula cabeza de serie. Para ello, se tenía hacer reaccionar la posición dada de esta molécula con un grupo amino primario, que contuviese el sustituyente R a probar.

Así pues, el trabajo consistía en seleccionar, de entre una base de datos de N aminas primarias reactivas, un número  $n$  de compuestos procurando que se mantuviera un alto porcentaje de la diversidad química disponible en la base de datos original. En el presente estudio, el tamaño de muestra  $n$  se fijó *a priori* en 50 compuestos por las características del estudio de química combinatoria para el que se realizó el estudio. La descripción química (descriptores moleculares) escogida para cada compuesto de la base de datos buscaba un compromiso entre calidad y velocidad, ya que el método tenía que manejar alrededor de 1000 compuestos. Así se optó por unos descriptores químico-cuánticos obtenidos a través de cálculos semiempíricos que incorporan información del entorno (solvente), proporcionando de esta forma información estérica, electrostática y de solvatación. El tratamiento estadístico posterior a través del análisis de componentes principales evita la redundancia en la información. La validez de estos descriptores ha sido evaluada en un estudio comparativo realizado en nuestro grupo (**apéndice II**), en el que el mismo conjunto de reactivos ha sido caracterizado con otros descriptores moleculares. La discusión y descripción de este estudio no es materia de la presente tesis, pero como se

puede comprobar en la referencia correspondiente,<sup>[64]</sup> la validez de los descriptores semiempíricos utilizados en el **Artículo I** ha quedado probada para este tipo de estudios. Finalmente, el análisis de conglomerados por el método de las k-medias permite no sólo la obtención de 50 representantes diferentes entre sí, que cubran toda la información contenida en la base de datos, sino la obtención de 50 grupos, cada uno ligado a cada representante. Esto hace posible la sustitución de un representante por otro compuesto de su mismo grupo cuando, por razones de síntesis, no sea posible la primera opción. Otra novedad en este trabajo es la definición de un coeficiente, denominado *R<sup>2</sup>-like* (tipo  $R^2$ ), capaz de medir la diversidad existente entre los compuestos representantes, con respecto a la diversidad total existente en la base de datos inicial, siempre dentro del espacio de descriptores propuesto. Este coeficiente ha sido empleado también con éxito en nuestro estudio posterior, lo que confirma su utilidad en estudios de diversidad molecular.

Dentro de los **métodos indirectos**, el campo donde nuestro grupo de investigación tiene mayor tradición es el de la semejanza molecular basada en campos de interacción molecular (ver **Sección 2.4** de la Introducción). A este respecto, en el marco del proyecto de colaboración en que se ubica la presente tesis, hemos realizado un estudio sobre antagonistas de receptores de adenosina presentado en el XII congreso de la Sociedad Española de Química Terapéutica (ver **apéndice III**). En ese trabajo, tres compuestos pertenecientes a familias químicas distintas han sido comparados y superpuestos en base a su potencial electrostático molecular con el programa MIPSIM, desarrollado en nuestro grupo de investigación. Los resultados muestran dos posibles modos de superposición entre las tres familias de compuestos, uno de los cuales ha servido como guía para la síntesis de nuevos 3-furfuril derivados de xantinas (datos no publicados). Actualmente, este estudio está sirviendo como base para otro más amplio, que incluye otras familias de antagonistas, y el estudio de éstas en base a potenciales de interacción molecular obtenidos con el programa GRID.

A lo largo del desarrollo del trabajo de esta tesis, en nuestro grupo de investigación hemos visto la necesidad de avanzar en la metodología que subyace a los estudios de semejanza molecular. Este hecho me ha llevado a tomar parte en distintos estudios sobre el tema, a pesar de no tener estos trabajos una aplicación directa en el campo de los receptores de adenosina. Los Artículos derivados de esta parte del trabajo se recogen en las referencias [33,63], cuyos resúmenes se muestran en los **apéndices IV y V**.



## 5. CONCLUSIONES

La presente tesis ha tratado de contribuir a una línea de investigación de descubrimiento de nuevos fármacos activos en receptores de adenosina por medio del desarrollo y aplicación de nuevas estrategias en el ámbito del “diseño de fármacos asistido por ordenador”. En el marco de esta amplia definición, y a lo largo de los más de cuatro años que ha durado mi periodo de formación doctoral, he intentado dar respuesta a algunos de los retos metodológicos particulares que se han planteado en relación con la caracterización estructural de los receptores de adenosina, las interacciones con sus ligandos y el diseño de éstos. De acuerdo con las prácticas habituales en investigación científica, buena parte de los resultados de mi trabajo de investigación durante el citado periodo se ha publicado en revistas de prestigio internacional y incluidas en los Journal Citation Reports. A continuación trataré de recapitular los avances que creo se han logrado en el presente trabajo:

- Hemos desarrollado una estrategia original para la selección de una muestra que cubra adecuadamente la diversidad molecular existente en una base de datos de compuestos químicos. Dicha estrategia consiste en el uso combinado de ciertos descriptores moleculares, análisis de componentes principales y análisis de conglomerados por el método *k-means*. Se ha definido una medida de diversidad molecular útil para monitorizar y comparar procesos de muestreo.
- Hemos construido un modelo de la región transmembrana del receptor A<sub>1</sub> humano de adenosina utilizando la información experimental disponible y métodos avanzados de modelado de GPCRs. El modelo obtenido es consistente con las evidencias experimentales y se ha demostrado útil para el estudio de la unión de ligandos.
- Para soslayar la incertidumbre experimental existente sobre la localización del sitio de unión de agonistas en el receptor considerado, se ha usado un mapeo del receptor con ribosa y una comparación del resultado con sitios de unión de ribosa conocidos experimentalmente, habiendo obtenido una solución consistente y única.
- Utilizando el modelo de receptor A<sub>1</sub> obtenido, hemos explorado y comparado el acoplamiento de una serie de agonistas, obteniendo un modo de unión que permite la formulación de relaciones estructura-actividad consistentes con información experimental.
- Asimismo, y a partir de los complejos agonista-receptor obtenidos se han llevado a cabo predicciones teóricas de las energías de *binding*, obteniendo un grado de acuerdo con los datos experimentales que resulta esperanzador si tenemos en cuenta la naturaleza no experimental de la geometría del receptor y la dificultad de modelado de los GPCRs.
- Consideramos que el protocolo propuesto y utilizado para la exploración y validación de modos de unión puede resultar una estrategia útil en situaciones análogas en las que se desconoce la geometría del complejo ligando-receptor.
- He colaborado en la línea de investigación que nuestro laboratorio mantiene sobre el desarrollo y aplicación de técnicas basadas en potenciales de interacción molecular para encontrar superposiciones no evidentes de moléculas con una actividad biológica. Concretamente la aplicación en el ámbito de los antagonistas de receptores de adenosina ha dado unos resultados preliminares prometedores en cuya investigación continuo implicado.



## 6. REFERENCIAS

1. Sánchez J. Farma 2010: el nuevo listón de la información. IBM consulting. <http://www-5.ibm.com/es/press/notas/2002/diciembre/farma.html>. accedido 11 diciembre 2003
2. Parra G, Blanco E, Guigo R. GeneID in Drosophila. *Genome Research* 2000;10:511-515.
3. Warshel A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. New York: Wiley VCH; 1991.
4. Bugrim A, Nikolskaya T, Nikolsky Y. Early prediction of drug metabolism and toxicity: systems biology approach and modeling. *Drug Discov Today* 2004;9:127 - 35.
5. Boobis A, Gundert-Remy U, Kremers P, Macheras P, Pelkonen O. In silico prediction of ADME and pharmacokinetics Report of an expert meeting organised by COST B15. *European Journal of Pharmaceutical Sciences* 2002;17:183 - 93.
6. McAlindon T, Formica M, Kabbara K, LaValley M, Lehmer M. Conducting clinical trials over the internet: feasibility study. *British Medical Journal* 2003;327:484 - 7.
7. Kubinyi H. *QSAR: Hansh analysis and related approaches*: VCH; 1993.
8. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001;291:1304-51.
9. Kubinyi H. Drug research: myths, hype and reality. *Nature Rev Drug Discov* 2003;2:665-668.
10. Liljefors T, Pettersson I. Computer-aided development of three dimensional pharmacofore models. In: P. Krosggaard-Larsen TLaUM, editor. *A Textbook of Drug Design and Development*; 1996.
11. Perutz MF. Structure of hemoglobin. *Brookhaven Symp Biol* 1960;13:165-83.
12. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 1958;181:662-6.
13. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, et al. Crystal structure of rhodopsin: A G protein-coupled receptor [see comments]. *Science* 2000;289:739-45.
14. Unger VM, Hargrave PA, Baldwin JM, Schertler GF. Arrangement of rhodopsin transmembrane alpha-helices. *Nature* 1997;389:203-6.
15. Schertler GF, Villa C, Henderson R. Projection structure of rhodopsin. *Nature* 1993;362:770-2.
16. Baldwin JM, Schertler GF, Unger VM. An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J Mol Biol* 1997;272:144-64.
17. Höltje H-D, Folkers G, Beier T. *Molecular modeling : basic principles and applications*. Weinheim ; New York: Vch; 1997.

18. Bourne PE, Weissig H. Structural bioinformatics. Hoboken, N.J.: Wiley-Liss; 2003.
19. Sternberg MJE. Protein structure prediction : a practical approach. Oxford ; New York: IRL Press at Oxford University Press; 1996.
20. Chivian D, Robertson T, Bonneau R, Baker D. Ab initio methods. In: Bourne PE, Weissig H, editors. Structural bioinformatics. Hoboken, N.J.: Wiley-Liss; 2003. p. xix, 649 , [32] of plates.
21. Anfinsen CB. Principles that govern the folding of protein chains. *Science* 1973;181:223-30.
22. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol* 1997;268:209-25.
23. Godzik A. Fold recognition methods. In: Bourne PE, Weissig H, editors. Structural bioinformatics. Hoboken, N.J.: Wiley-Liss; 2003. p. xix, 649 , [32] of plates.
24. Rychlewski L, Jaroszewski L, Li W, Godzik A. Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci* 2000;9:232-41.
25. Sippl MJ, Weitckus S. Detection of native-like models for amino acid sequences of unknown three-dimensional structure in a data base of known protein conformations. *Proteins* 1992;13:258-71.
26. Jones DT, Taylor WR, Thornton JM. A new approach to protein fold recognition. *Nature* 1992;358:86-9.
27. Rost B. Twilight zone of protein sequence alignments. *Protein Eng* 1999;12:85-94.
28. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779-815.
29. Lattman E. Protein structure prediction: a special issue. *Proteins* 1995;23:295-460.
30. Von Heijne G. Prediction of transmembrane protein topology. In: Sternberg MJE, editor. Protein structure prediction : a practical approach. Oxford ; New York: IRL Press at Oxford University Press; 1996. p. 101-108.
31. Ballesteros JA, Weinstein H. Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors. In: Conn PM, Sealfon SC, editors. *Methods in Neurosciences*. San Diego: Academic Press; 1994. p. 366-428.
32. Carrieri A, Centeno NB, Rodrigo J, Sanz F, Carotti A. Theoretical evidence of a salt bridge disruption as the initiating process for the  $\alpha_{1d}$ -adrenergic receptor activation: a molecular dynamics and docking study. *Proteins* 2001;43:382-394.
33. Barbany M, Gutierrez de Teran H, Sanz F, Villà-Freixa J. Towards a MIP-based alignment and docking in computer-aided drug design. *Proteins* 2004;in press.
34. McMartin C, Bohacek RS. QXP: powerful, rapid computer algorithms for structure-based drug design. *J Comput Aided Mol Des* 1997;11:333-44.
35. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* 2002;47:409-43.

36. Klauben H, Buning C, Rarey M, Lengauer T. FlexE: Efficient molecular docking considering protein structure variations. *J Mol Biol* 2001;308:377-395.
37. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* 2002;46:34-40.
38. Lyne PD. Structure-based virtual screening: an overview. *Drug Discov Today* 2002;7:1047-55.
39. Perez C, Ortiz AR. Evaluation of docking functions for protein-ligand docking. *J Med Chem* 2001;44:3768-85.
40. Goodford PJ. A computational procedure for determining energetically favorable binding sites of biologically important macromolecules. *J Med Chem* 1985;28:849-857.
41. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* 1995;117:5179-5197.
42. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 1996;261:470-489.
43. Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem* 1999;42:791-804.
44. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* 2000;295:337-56.
45. Zwanzig RW. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J Chem Phys* 1954;22:1420.
46. Srinivasan JC, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J Am Chem Soc* 1998;120:9401-9409.
47. Åqvist J, Medina C, Samuelsson JE. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng* 1994;7:385-91.
48. Lee FS, Warshel A. A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. *J Chem Phys* 1992;97:3100-3107.
49. Hansson T, Marelus J, Åqvist J. Ligand binding affinity prediction by linear interaction energy methods. *J Comput Aided Mol Des* 1998;12:27-35.
50. Åqvist J, Hansson T. On the Validity of Electrostatic Linear Response in Polar Solvents. *J Phys Chem* 1996;100:9512-9521.
51. Hansch C, Fujita T. Rho-Sigma-Pi Analysis. Method for Correlation of Biological Activity and Chemical Structure. *Journal of the American Chemical Society* 1964;86:1616.
52. Hammett LP. *Physical organic chemistry; reaction rates, equilibria, and mechanisms*. 2nd ed. New York, London,: McGraw-Hill; 1970.
53. Free SM, Wilson JW. Mathematical Contribution to Structure-Activity Studies. *J Med Chem* 1964;7:395.

54. Cramer RD, Patterson DE, Bunce JD. Comparative Molecular-Field Analysis (Comfa) .1. Effect of Shape on Binding of Steroids to Carrier Proteins. *Journal of the American Chemical Society* 1988;110:5959-5967.
55. Baroni M, Costantino G, Cruciani G, Riganelli D, Valigi R, Clementi S. Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct. -Act. Relat.* 1993;12:9-20.
56. Pastor M, Cruciani G, Clementi S. Smart region definition: a new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships. *J Med Chem* 1997;40:1455-1464.
57. Baroni M, Clemanti S, Cruciani G, Kettaneh-Wold N, Wold S. D-Optimal design in QSAR. *Quant Struct-Act Relat* 1993;12:225-231.
58. Krosggaard-Larsen P, Liljefors T, Madsen U. *Textbook of drug design and discovery*. 3rd ed. London ; New York: Taylor & Francis; 2002.
59. Pastor M, Cruciani G, McLay I, Pickett S, Clementi S. GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J Med Chem* 2000;43:3233-43.
60. De Caceres M, Villà J, Lozano JJ, Sanz F. MIPSIM: similarity analysis of molecular interaction potentials. *Bioinformatics* 2000;16:568-569.
61. Bagdassarian CK, Schramm VL, Schwartz SD. Molecular Electrostatic Potential Analysis for Enzymatic Substrates, Competitive Inhibitors, and Transition-State Inhibitors. *J Am Chem Soc* 1996;118:8825-8836.
62. Hodgkin E, Richards W. Molecular Similarity Based on Electrostatic Potential and Electric Field. *Int J Quan Chem* 1987;14:105-110.
63. Barbany M, Gutierrez-de-Teran H, Sanz F, Villà-Freixa J, Warshel A. On the Generation of Catalytic Antibodies by Transition State Analogues. *ChemBioChem* 2003;4:277 - 85.
64. Fontaine F, Pastor M, Gutierrez de Teran H, Lozano JJ, Sanz F. Use of alignment-free molecular descriptors in diversity analysis and optimal sampling of molecular libraries. *Mol Diversity* 2003;6:135-147.
65. Car R, Parrinello M. Unified Approach for Molecular-Dynamics and Density-Functional Theory. *Phys Rev Lett* 1985;55:2471-2474.
66. Gaussian 98. Gaussian, Inc. Pittsburgh, PA, USA; 1998.
67. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, J. H. Gordon MSJ, Koseki S, et al. GAMESS. *J. Comput. Chem.* 1993;14:1347-1363.
68. Dewar MJS, Thiel W. Ground States for molecules. 38. The MNDO method. Aproximations and parameters. *J Am Chem Soc* 1977;99:4907-4917.
69. Dewar MJS, Zoebish EG, E.F. H, Stewart JJ. AM1: A new general purpose Quantum Mechanical model. *J Am Chem Soc* 1985;107:3902-3909.
70. Stewart JJ. MOPAC: a semiempirical molecular orbital program. *J Comput Aided Mol Des* 1990;4:1-105.
71. Hawkins GD, Giesen DJ, G.C. L, Chambers CC, Rossi I, Storer JW, et al. AMSOL. In. 6.5.2 ed. Minneapolis: Univeristy of Minnesota; 1997.

72. Singh UC, Kollman PA. A Combined Ab Initio Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the  $\text{CH}_3\text{Cl} + \text{Cl}^-$  Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* 1986;7:718-730.
73. Gao J. Hybrid Quantum and Molecular Mechanical Simulations: an Alternative Avenue to Solvent Effects in Organic Chemistry. *Acc Chem Res* 1996;29:298-305.
74. Bash PA, Field MJ, Davenport RC, Petsko GA, Ringe D, Karplus M. Computer Simulation and Analysis of the Reaction Pathway of Triosephosphate Isomerase. *Biochemistry* 1991;30:5826-5832.
75. Warshel A, Levitt M. Theoretical Studies of Enzymatic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J Mol Biol* 1976;103:227-249.
76. Kicska GA, Tyler PC, Evans GB, Furneaux RH, Kim K, Schramm VL. Transition State Analogue Inhibitors of Purine Nucleoside Phosphorylase from *Plasmodium falciparum*. *J Biol Chem* 2002;277:3219-3225.
77. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity-Rapid acces to atomic charges. *Tetrahedron* 1980;36:3219-3228.
78. Bayly CI, Cieplak P, Cornell WD, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* 1993;97:10269-10280.
79. Warshel A. What about protein polarity? *Nature* 1987;330:15-16.
80. Schutz CN WA. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* 2001;44:400-417.
81. Hopfinger AJ. Conformational properties of macromolecules. New York,: Academic Press; 1973.
82. Hünenberger P, van Gunsteren W. Empirical classic force fields for molecular systems. In: *Arbeitstagung für Theoretische Chemie*; 1996; Österreich; 1996.
83. Case DA, Pearlman DA, Caldwell JW, Cheatham III TE, Ross WS, Simmerling CL, et al. AMBER. In. 6 ed. San Francisco: University of California; 1999.
84. Marelus J, Kolmodin K, Feierberg I, Åqvist J. Q: An MD program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J Mol Graph Modelling* 1999;16:213-225.
85. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 1998;19:1639-1662.
86. Verlet L. Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys Rev* 1967;159(1):98.
87. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys* 1984;81:3684-3690.
88. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equation of motion of a system constraints: molecular dynamics of N-alkanes. *J Compt Phys* 1977;23:327-341.

89. Miller M, Wales D. Energy landscape of a model protein. *J Chem Phys* 1999;111:6610-6616.
90. Levitt M. A simplified representation of protein conformations for rapid simulation of protein folding. *J Mol Biol* 1976;104:59-107.
91. Wold S, Esbensen K, Geladi P. Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems* 1987;2:37-52.
92. Joliffe IT, Morgan BJT. Principal component analysis and exploratory factor analysis. *Stat Meth Med Res* 1992;1:69-95.
93. Eriksson L, Johansson E. Multivariate design and modeling in QSAR. *Chemometrics and Intelligent Laboratory Systems* 1996;34:1-19.
94. Downs GM, Barnard JM. Clustering methods and their uses in computational chemistry. *Reviews in Computational Chemistry*, Vol 18 2002:1-40.
95. Vogt W, Nagel D, Sator H. Cluster analysis in clinical chemistry : a model. Chichester ; New York: Wiley; 1987.
96. Marengo E, Todeschini R. A new algorithm for optimal, distance-based experimental design. *Chemom Intell Lab Sys* 1992;16:37-44.
97. Fredholm BB, IJzerman AP, Jacobson KA, Klotz KN, Linden J. International Union of Pharmacology. XXV. Nomenclature and classification of adenosine receptors. *Pharmacol Rev* 2001;53:527-52.
98. Lee CH, Jiang M, Cowart M, Gfesser G, Perner R, Kim KH, et al. Discovery of 4-amino-5-(3-bromophenyl)-7-(6-morpholino-pyridin-3-yl)pyrido[2,3-d]pyrimidine, an orally active, non-nucleoside adenosine kinase inhibitor. *J Med Chem* 2001;44:2133-8.
99. Müller CE. Adenosine receptor ligands-recent developments part I. Agonists. *Curr Med Chem* 2000;7:1269-88.
100. George SR, O'Dowd BF, Lee SR. G-protein-coupled receptor oligomerization and its potential for drug discovery. *Nature Rev Drug Discov* 2002;1:808-820.
101. Bockaert J. G Protein-coupled Receptors. In: *Encyclopedia of Life Sciences*. London: Nature Publishing Group; 1999.
102. Ji TH, Grossmann M, Ji I. G Protein-coupled Receptors. I. Diversity of receptor-Ligand interactions. *J Biol Chem* 1998;273:17299-17302.
103. Horn F, Weare J, Beukers M, Horsch S, Bairoch A, Chen W, et al. GPCRDB: an information system for G protein-coupled receptors. *Nucl. Acids. Res.* 1998;26:275-279.
104. Hibert M, Trumpp-Kallmeyer S, Bruinvels A, Hoflack J. Three-dimensional models of neurotransmitter G-binding protein-coupled receptors. *Mol Pharmacol* 1991;40:8-15.
105. Pogozheva ID, Lomize AL, Mosberg HI. The transmembrane 7- $\alpha$ -bundle of rhodopsin: distance geometry calculations with hydrogen bonding constraints. *Biophys J* 1997;70:1963-1985.

106. Colquhoun D. Binding, gating, affinity and efficacy: the interpretation of structure-activity relationships for agonists and of the effects of mutating receptors. *Br J Pharmacol* 1998;125:924-947.
107. Flower DR. Modelling G-protein-coupled receptors for drug design. *Biochimica et Biophysica Acta* 1999;1422:207-34.
108. Kristiansen K, Dahl SG, Edvardsen O. A database of mutants and effects of site-directed mutagenesis experiments on G protein-coupled receptors. *Proteins-Structure Function and Genetics* 1996;26:81-94.
109. Wang C, Gallaher T, Shih J. Site-directed mutagenesis of the serotonin 5-hydroxytryptamine<sub>2</sub> receptor: identification of amino acids necessary for ligand binding and receptor activation. *Mol Pharmacol* 1993;43:931-940.
110. Kristiansen K, Kroeze WK, Willins DL, Gelber EI, Savage JE, Glennon RA, et al. A Highly Conserved Aspartic Acid (Asp-155) Anchors the Terminal Amine Moiety of Tryptamines and Is Involved in Membrane Targeting of the 5-HT<sub>2A</sub> Serotonin Receptor But Does Not Participate in Activation via a "Salt-Bridge Disruption" Mechanism. *J Pharmacol Exp Ther* 2000;293:735-746.
111. Scheer A, Fanelli F, Costa T, De Benedetti P, Cotecchia S. Constitutively active mutants of the alpha 1B-adrenergic receptor: role of highly conserved polar amino acids in receptor activation. *EMBO J* 1996;15:3566-3578.
112. Zhou W, Flanagan C, Ballesteros J, Konvicka K, Davidson J, Weinstein H, et al. A reciprocal mutation supports helix 2 and helix 7 proximity in the gonadotropin-releasing hormone receptor. *Mol Pharmacol* 1994;45:165-170.
113. Altenbach C, Klein-Seetharaman J, Cai KW, Khorana HG, Hubbell WL. Structure and function in rhodopsin: Mapping light-dependent changes in distance between residue 316 in helix 8 and residues in the sequence 60-75, covering the cytoplasmic end of helices TM1 and TM2 and their connection loop CL1. *Biochemistry* 2001;40:15493-15500.
114. Altenbach C, Cai KW, Klein-Seetharaman J, Khorana FG, Hubbell WL. Structure and function in rhodopsin: Mapping light-dependent changes in distance between residue 65 in helix TM1 and residues in the sequence 306-319 at the cytoplasmic end of helix TM7 and in helix H8. *Biochemistry* 2001;40:15483-15492.
115. Dawson ES, Wells JN. Determination of Amino Acid Residues that are Accessible from the Ligand Binding Crevice in the Seventh Transmembrane-Spanning Region of the Human A<sub>1</sub> Adenosine Receptor. *Mol Pharmacol* 2001;59:1187-1195.
116. Kennedy A, Mangum K, Linden J, Wells J. Covalent modification of transmembrane span III of the A<sub>1</sub> adenosine receptor with an antagonist photoaffinity probe. *Mol Pharmacol* 1996;50:789-798.
117. Mirzadegan T, Benko G, Filipek S, Palczewski K. Sequence analyses of G-protein-coupled receptors: Similarities to rhodopsin. *Biochemistry* 2003;42:2759-2767.
118. Goodman LS, Gilman A, Hardman JG, Gilman AG, Limbird LE, Molinoff PB, et al. Goodman & Gilman's the pharmacological basis of therapeutics. 9th ed. New York: McGraw-Hill Health Professions Division; 1996.
119. Horn F, van der Wenden EM, Oliveira L, AP IJ, Vriend G. Receptors coupling to G proteins: is there a signal behind the sequence? *Proteins* 2000;41:448-59.

120. Samama P, Cotecchia S, Costa T, Lefkowitz RJ. A mutation-induced activated state of the beta 2-adrenergic receptor. Extending the ternary complex model. *J Biol Chem* 1993;268:4625-36.
121. Kenakin T. Efficacy at G-protein-coupled receptors. *Nature Rev Drug Discov* 2002;1:103-110.
122. Hunyady L, Vauquelin G, Vanderheyden P. Agonist induction and conformational selection during activation of a G-protein-coupled receptor. *Trends Pharmacol Sci* 2003;24:81-86.
123. Drury AN, Szent-Gyorgyi A. The physiological activity of adenine compounds with especial reference to their action upon the mammalian heart. *J Physiol. (Lond)* 1929;68:213-237.
124. Burnstock G. Cell membrane receptors for drugs and hormones: a multidisciplinary approach. New York: Raven Press; 1978.
125. Williams M, Jarvis MF. Purinergic and pyrimidinergic receptors as potential drug targets. *Biochem Pharmacol* 2000;59:1173-1185.
126. Müller C, Stein B. Adenosine Receptor Antagonists: Structures and Potential Therapeutic Applications. *Curr Pharm Design* 1996;2:501-530.
127. Fredholm BB, Arslan G, Kull B, Svenningsson P. Locating the neuronal targets for caffeine. *Drug Dev Res* 1998;45:324-328.
128. Fredholm BB, Arslan G, Halldner L, Kull B, Schulte G, Wasserman W. Structure and function of adenosine receptors and their genes. *Naunyn Schmiedebergs Arch Pharmacol* 2000;362:364-74.
129. Haas HL, Selbach O. Functions of neuronal adenosine receptors. *Naunyn Schmiedebergs Arch Pharmacol* 2000;362:375-81.
130. Kaiser SM, Quinn RJ. Adenosine receptors as potential therapeutic targets. *Drug Discov Today* 1999;4:542-551.
131. Franco R, Ferre S, Torvinen M, Gines S, Hillion J, Ciruela F, et al. Adenosine/dopamine receptor-receptor interactions in the central nervous system. *Drug Dev Res* 2001;52:296-302.
132. Richardson PJ, Kase H, Jenner PG. Adenosine A(2A) receptor antagonists as new agents for the treatment of Parkinson's disease. *Trends Pharmacol Sci* 1997;18:338-344.
133. Ciruela F, Escriche M, Soloviev MM, Canela EI, Burgeno J, Mallol J, et al. Adenosine-glutamate receptor-receptor interactions in the central nervous system. *Drug Dev Res* 2001;52:316-322.
134. Bertolet B, Hill J, Kerensky R, Belardinelli L. Myocardial infarction related atrial fibrillation: role of endogenous adenosine. *Heart* 1997;78:88-90.
135. Neely CF, DiPierro FV, Kong M, Greelish JP, Gardner TJ. A<sub>1</sub> adenosine receptor antagonists block ischemia-reperfusion injury of the heart. *Circulation* 1996;94:II376-80.
136. Headrick JP, Hack B, Ashton KJ. Acute adenosinergic cardioprotection in ischemic-reperfused hearts. *Am J Physiol Heart Circ Physiol* 2003;285:H1797-1818.

137. Palmer TM, Stiles GL. Adenosine Receptors. *Neuropharmacology* 1995;34:683-694.
138. Olsson RA, Pearson JD. Cardiovascular Purinoceptors. *Physiol Rev* 1990;70:761-845.
139. Suzuki F, Shimada J, Mizumoto H, Karasawa A, Kubo K, Nonaka H, et al. Adenosine A<sub>1</sub> antagonists. 2. Structure-activity relationships on diuretic activities and protective effects against acute renal failure. *J Med Chem* 1992;35:3066-75.
140. Szabo C, Scott G, Virag L, Egnaczyk G, Salzman A, Shanley T, et al. Suppression of macrophage inflammatory protein (MIP)-1 $\alpha$  production and collagen-induced arthritis by adenosine receptor agonists. *Br J Pharmacol* 1998;125:379-387.
141. Cronstein BN. The antirheumatic agents sulphasalazine and methotrexate share an anti-inflammatory mechanism. *Br J Rheumatol* 1995;34:30-32.
142. Ohta A, Sitkovsky M. Role of G-protein-coupled adenosine receptors in downregulation of inflammation and protection from tissue damage. *Nature* 2001;414:916-920.
143. Rosengren S, Bong GW, Firestein GS. Antiinflammatory Effects of an Adenosine Kinase Inhibitor - Decreased Neutrophil Accumulation and Vascular Leakage. *J Immunol* 1995;154:5444-5451.
144. Feoktistov I, Polosa R, Holgate ST, Biaggioni I. Adenosine A<sub>2B</sub> receptors: a novel therapeutic target in asthma? *Trends Pharmacol Sci* 1998;19:148-153.
145. Baraldi PG, Cacciari B, Romagnoli R, Merighi S, Varani K, Borea PA, et al. A<sub>3</sub> adenosine receptor ligands: History and perspectives. *Med Res Rev* 2000;20:103-128.
146. Kim Y-C, Ji X-D, Melman N, Linden J, Jacobson KA. Anilinde derivatives of an 8-phenylxanthine carboxylic congener are highly potent and selective antagonist at human A<sub>2B</sub> adenosine receptors. *J Med Chem* 2000;43:1165-1172.
147. Fan M, Qin WX, Mustafa SJ. Characterization of adenosine receptor(s) involved in adenosine-induced bronchoconstriction in an allergic mouse model. *Am J Physiol Lung Cell Mol Physiol* 2003;284:L1012-L1019.
148. Vidal B, Esteve C, Segarra V, Raviña E, Fernandez F, Loza MI, Sanz, F., inventors; Almirall Prodesfarma, assignee. 6-Phenyldihydropyrrolopyrimidinedione derivatives. Spain patent WO 03/000694. 2003 3 January 2003.
149. Olah ME, Stiles GL. The role of receptor structure in determining adenosine receptor activity. *Pharmacol Ther* 2000;85:55-75.
150. Ijzerman A, van Galen PJM, Jacobson KA. Molecular modeling of adenosine receptors. I. The ligand binding site on the A<sub>1</sub> receptor. *Drug Des Discov* 1992;9:49-67.
151. Bianucci A-M, Bigi MU, Biagi G, Giorgi I, Livi O, Scartoni V. A 3D Model of the Human A<sub>1</sub> Adenosine Receptor. An Evaluation of the Binding Free-Energy with Ligands. *Drug Des Discov* 1998;15:149-156.
152. Rivkees SA, Barbhuiya H, Ijzerman AP. Identification of the Adenine Binding Site of the Human A<sub>1</sub> Adenosine Receptor. *J Biol Chem* 1999;274:3617-3621.

153. Da Settimo F, Primofiore G, Taliani S, Marini AM, Motta CL, Novellino E, et al. 3-Aryl[1,2,4]triazolo[4,3-a]benzimidazol-4(10H)-ones: A New Class of Selective A<sub>1</sub> Adenosine Receptor Antagonists. *J Med Chem* 2001;44:316-327.
154. Bondavalli F, Botta M, Bruno O, Ciacci A, Corelli F, Fossa P, et al. Synthesis, molecular modeling studies, and pharmacological activity of selective A(1) receptor antagonists. *J Med Chem* 2002;45:4875-4887.
155. Ijzerman A, van der Wenden EM, van Galen PJM, Jacobson KA. Molecular modelling of adenosine receptors. The ligand binding site on the rat adenosine A<sub>2A</sub> receptor. *Eur J Pharmacol* 1994;268:95-104.
156. Kim J, Wess J, van Rhee AM, Schöneberg T, Jacobson KA. Site-directed Mutagenesis Identifies Residues Involved in Ligand Recognition in the Human A<sub>2A</sub> Adenosine Receptor. *J Biol Chem* 1995;270:13987-13997.
157. Kim S-K, Gao Z-G, Rompaey PV, Gross AS, Chen A, Calenbergh SV, Jacobson, KA. Modeling the Adenosine Receptors: Comparison of the Binding Domains of A<sub>2A</sub> Agonists and Antagonists. *J Med Chem* 2003;46:4847-59.
158. Pastorin G, Da Ros T, Spalluto G, Deflorian F, Moro S, Cacciari B, et al. Pyrazolo[4,3-e]-1,2,4-triazolo[1,5-c]pyrimidine derivatives as adenosine receptor antagonists. Influence of the N5 substituent on the affinity at the human A<sub>3</sub> and A<sub>2B</sub> adenosine receptor subtypes: A molecular modeling investigation. *J Med Chem* 2003;46:4287-4296.
159. Li AH, Moro S, Forsyth N, Melman N, Ji XD, Jacobson KA. Synthesis, CoMFA analysis, and receptor docking of 3,5-diacyl-2,4-dialkylpyridine derivatives as selective A<sub>3</sub> adenosine receptor antagonists. *J Med Chem* 1999;42:706-721.
160. Baraldi PG, Cacciari B, Moro S, Romagnoli R, Ji X, Jacobson KA, et al. Fluorosulfonyl- and bis-(beta-chloroethyl)amino-phenylamino functionalized pyrazolo[4,3-e]1,2,4-triazolo[1,5-c]pyrimidine derivatives: irreversible antagonists at the human A<sub>3</sub> adenosine receptor and molecular modeling studies. *J Med Chem* 2001;44:2735-42.
161. Jacobson KA, Gao ZG, Chen A, Barak D, Kim SA, Lee K, et al. Neoreceptor concept based on molecular complementarity in GPCRs: a mutant adenosine A(3) receptor with selectively enhanced affinity for amine-modified nucleosides. *J Med Chem* 2001;44:4125-36.
162. Cristalli G, Costanzi S, Lambertucci C, Taffi S, Vittori S, Volpini R. Purine and deazapurine nucleosides: synthetic approaches, molecular modelling and biological activity. *Farmaco* 2003;58:193-204.
163. Barbhaiya H, McClain R, Ijzerman A, Rivkees SA. Site-directed mutagenesis of the human A<sub>1</sub> adenosine receptor: influences of acidic and hydroxy residues in the first four transmembrane domains on ligand binding. *Mol Pharmacol* 1996;50:1635-42.
164. Scholl DJ, Wells JN. Serine and alanine mutagenesis of the nine native cysteine residues of the human A<sub>1</sub> adenosine receptor. *Biochem Pharmacol* 2000;60:1647-54.
165. Olah ME, Ren H, Ostrowski J, Jacobson KA, Stiles GL. Cloning, expression, and characterization of the unique bovine A<sub>1</sub> adenosine receptor. Studies on the ligand binding site by site-directed mutagenesis. *J Biol Chem* 1992;267:10764-70.

166. Tucker AL, Robeva AS, Taylor HE, Holeton D, Bockner M, Lynch KR, et al. A<sub>1</sub> adenosine receptors. Two amino acids are responsible for species differences in ligand recognition. *J Biol Chem* 1994;269:27900-6.
167. Townsend-Nicholson A, Schofield PR. A threonine residue in the seventh transmembrane domain of the human A<sub>1</sub> adenosine receptor mediates specific agonist binding. *J Biol Chem* 1994;269:2373-6.
168. Klotz KN. Adenosine receptors and their ligands. *Naunyn Schmiedebergs Arch Pharmacol* 2000;362:382-391.
169. Müller CE. A<sub>1</sub> adenosine receptors and their ligands: overview and recent developments. *Farmacologia* 2001;56:77-80.
170. Cristalli G, Volpini R. Preface. *Curr Top Med Chem* 2003;3 (issue 4).
171. Lohse MJ, Klotz KN, Schwabe U, Cristalli G, Vittori S, Grifantini M. 2-Chloro-N6-cyclopentyladenosine: a highly selective agonist at A<sub>1</sub> adenosine receptors. *Naunyn Schmiedebergs Arch Pharmacol* 1988;337:687-9.
172. Daly JW, Padgett WL. Agonist activity of 2- and 5'-substituted adenosine analogs and their N6-cycloalkyl derivatives at A<sub>1</sub> and A<sub>2</sub> adenosine receptors coupled to adenylate cyclase. *Biochem Pharmacol* 1992;43:1089-93.
173. Van der Wenden EMC, M.; Roelen, H.; Lorenzen, A.; Von Frijtag Drabbe künzel, J. K.; IJzerman, A. 5'-Substituted adenosine-analogues as new high-affinity partial agonists for the adenosine A<sub>1</sub> receptor. *J Med Chem* 1998;41:102-108.
174. Cristalli GG, M.; Vittori, S.; Balduini, W.; Cattabeni, F. Adenosine and 2-chloroadenosine deaza-analogues as adenosine receptor agonists. *Nucleoside Nucleotides* 1985;4:625-639.
175. Vittori S, Lorenzen A, Stannek C, Costanzi S, Volpini R, AP IJ, et al. N-cycloalkyl derivatives of adenosine and 1-deazaadenosine as agonists and partial agonists of the A<sub>1</sub> adenosine receptor. *J Med Chem* 2000;43:250-60.
176. van der Wenden EM, von Frijtag Drabbe Kunzel JK, Mathot RA, Danhof M, AP IJ, Soudijn W. Ribose-modified adenosine analogues as potential partial agonists for the adenosine receptor. *J Med Chem* 1995;38:4000-6.
177. Lohse MJ, Klotz KN, Diekmann E, Friedrich K, Schwabe U. 2',3'-Dideoxy-N6-cyclohexyladenosine: an adenosine derivative with antagonist properties at adenosine receptors. *Eur J Pharmacol* 1988;156:157-60.
178. Jacobson KA, Ji XD, Li AH, Melman N, Siddiqui MA, Shin KJ, et al. Methanocarba analogues of purine nucleosides as potent and selective adenosine receptor agonists. *J Med Chem* 2000;43:2196-2203.
179. Jacobson KA, Gallo-Rodriguez C, Melman N, Fischer B, Maillard M, Van Bergen A, et al. Structure-activity relationships of 8-stylylxanthines as A<sub>2</sub>-selective adenosine antagonists. *J Med Chem* 1993;36:1333-1342.
180. Van der Wenden EM, IJzerman AP, Soudijn W. A steric and electrostatic comparison of three models for the agonist/antagonist binding site on the adenosine A<sub>1</sub> receptor. *J Med Chem* 1992;35:629-635.

181. Peet NP, Lentz NL, Meng EC, Dudley MW, Ogden AM, Demeter DA, et al. A novel synthesis of xanthines: support for a new binding mode for xanthines with respect to adenosine at adenosine receptors. *J Med Chem* 1990;33:3127-30.
182. Rodrigo J, Barbany M, Gutierrez de Teran H, Centeno NB, de Caceres M, Dezi C, et al. Comparison of Biomolecules on the Basis of Molecular Interaction Potentials. *J Brazil Chem Soc* 2002;13:795-9.
183. Grahner B, Winiwarter S, Lanzner W, Muller CE. Synthesis and structure-activity relationships of deazaxanthines: analogs of potent A<sub>1</sub> and A<sub>2</sub> adenosine receptor antagonists. *J Med Chem* 1994;37:1526-34.
184. Francis JE, Cash WD, Psychoyos S, Ghai G, Wenk P, Friedmann RC, et al. Structure Activity Profile of a Series of Novel Triazoloquinazoline Adenosine Antagonists. *J Med Chem* 1988;31:1014-1020.
185. Palmer T, Poucher S, Jacobson K, Stiles G. 125I-4-(2-[7-amino-2-[2-furyl][1,2,4]triazolo[2,3-a][1,3,5] triazin-5-yl-amino]ethyl)phenol, a high affinity antagonist radioligand selective for the A<sub>2a</sub> adenosine receptor. *Mol Pharmacol* 1995;48:970-974.
186. Baraldi PG, Cacciari B, Spalluto G, Villatoro M, Zocchi C, Dionisotti S, et al. Pyrazolo[4,3-e]-1,2,4-triazolo[1,5-c]pyrimidine derivatives: Potent and selective A<sub>2A</sub> adenosine antagonists. *J Med Chem* 1996;39:1164-1171.
187. Müller CE, Geis U, Grahner B, Lanzner W, Eger K. Chiral pyrrolo [2,3-d]pyrimidine and pyrimido[4,5-b]indole derivatives: Structure-activity relationships of potent, highly stereoselective A<sub>1</sub>-adenosine receptor antagonists. *J Med Chem* 1996;39:2482-2491.
188. Ferrarini PL, Mori C, Manera C, Martinelli A, Mori F, Saccomanni G, et al. A novel class of highly potent and selective A<sub>1</sub> adenosine antagonists: Structure-affinity profile of a series of 1,8-naphthyridine derivatives. *J Med Chem* 2000;43:2814-2823.
189. Moro S, Li AH, Jacobson KA. Molecular modeling studies of human A<sub>3</sub> adenosine antagonists: structural homology and receptor docking. *J Chem Inf Comput Sci* 1998;38:1239-48.
190. IsisBase2.3. MDL information systems Inc; 1990-2000.
191. Brea J, Rodrigo J, Carrieri A, Sanz F, Cadavid MI, Enguix MJ, et al. New serotonin 5-HT<sub>2A</sub>, 5-HT<sub>2B</sub>, and 5HT<sub>2C</sub> receptor antagonists: synthesis, pharmacology, 3D-QSAR, and molecular modeling of (aminoalkyl) benzo and heterocycloalkanones. *J Med Chem* 2002;45:54-71.

## 7. ARTÍCULOS

Esta Tesis se basa en los siguientes artículos, numerados por números romanos tal y como se viene indicando en el texto:

- I. Gutiérrez-de-Terán, H.; Lozano, J.J.; Segarra, V.; Sanz, F. Molecular diversity strategy on the basis of quantum mechanical computations and principal component analysis; *Comb. Chem. High T. Scr.* **2002**, 5: 39-47.
- II. Gutiérrez-de-Terán, H.; Centeno, N.B.; Pastor, M.; Sanz, F. Novel Approaches for modelling of the A<sub>1</sub> adenosine receptor and its agonist binding site. *Proteins* **2004**, 54:705-715
- III. Gutiérrez-de-Terán, H.; Pastor, M.; Centeno, N.B.; Åqvist, J.; Sanz, F. Comparative analysis of putative agonist binding modes in the human A<sub>1</sub> adenosine receptor. *ChemBioChem*, (in press)

Los siguientes artículos, si bien no son parte formal de los resultados de esta tesis, son el resultado de colaboraciones mantenidas en el periodo de formación doctoral. El resumen de algunos de ellos se muestra como apéndice:

- Barbany, M.\*; Gutiérrez-de-Terán, H.\*; Sanz, F.; Villà-Freixa, J.; Towards a MIP-based alignment and docking in computer-aided drug design. *Proteins* (in press<sup>s</sup>)
- Fontaine, F.; Pastor, M.; Gutiérrez-de-Terán, H.; Lozano, J.J. Sanz, F. Use of alignment-free molecular descriptors in diversity analysis and optimal sampling of molecular libraries. *Mol. Divers.* **2003**, 6:135-147
- Barbany, M.; Gutiérrez-de-Terán, H.; Sanz, F.; Villà-Freixa, J.; Warshel, A. On the generation of catalytic antibodies by transition state analogs; *ChemBioChem* **2003**, 4: 277-285.
- Rodrigo, J.; Barbany, M.; Gutiérrez-de-Terán, H.; Centeno, N.B.; de-Cáceres, M.; Dezi, C.; Fontaine, F.; Lozano, J.J.; Pastor, M.; Villà, J.; Sanz, F: Comparison of biomolecules on the basis of molecular interaction potentials; *J. Brazil. Chem. Soc.* **2002**, 13: 795-799.

\* Los dos autores comparten la primera autoría



## **ARTÍCULO I**



# Molecular Diversity Sample Generation on the Basis of Quantum-Mechanical Computations and Principal Component Analysis

H. Gutiérrez-de-Terán<sup>a,†</sup>, J. J. Lozano<sup>a,#,†</sup>, V. Segarra<sup>b</sup> and F. Sanz<sup>a,\*</sup>

<sup>a</sup>Research Group on Biomedical Informatics, IMIM, Universitat Pompeu Fabra, C/ Dr. Aiguader 80, E-08003 Barcelona, Spain

<sup>b</sup>Almirall Prodesfarma, S.A., C/ Cardener 68-74, E-08024 Barcelona, Spain

**Abstract:** The present study introduces a new strategy of selection of a maximum diversity sample of  $n$  compounds from  $N$  available in a molecular database. This strategy can be useful in pharmacological screening, combinatorial chemistry or parallel synthesis planning. It consists of first describing the compounds by means of parameters derived from quantum mechanical computations (water solvation  $G$ , benzene solvation  $G$ , octanol solvation  $G$ , dipolar moment), as well as standard molecular parameters such as solvent-accessible surface area and molecular weight. Solvation parameters are used because of the importance of this phenomenon in the pharmacological behaviour. Redundant information in the description of the compounds is eliminated by using principal components (PC) instead of the original descriptors. Based on the similarity between the  $N$  compounds in the PC space, they are classified into  $n$  groups by  $k$ -means cluster analysis. The compounds that are nearest to the centroid of each cluster constituted the maximum diversity sample. When practical difficulties exist for the use of one of the proposed compounds, another also close to the cluster centroid can substitute for it. This strategy has been tested in the selection of a sample of 50 amines from the 923 available in the Aldrich catalogue. The results have been contrasted with those obtained from an optimal, distance-based experimental design, resulting in an 86% of agreement between both approaches. An  $R^2$ -like diversity coefficient has been used to assess the quality of the proposed solutions.

## INTRODUCTION

The selection of an assorted sample of compounds from those available in a molecular database is a typical task in pharmacological screening, combinatorial chemistry or parallel synthesis planning. For instance, in the framework of a lead optimization process, it is a common practice to systematically explore a certain position of a lead compound by testing the introduction of a series of substituents having diverse molecular properties. The introduction of combinatorial chemistry techniques has extended and accelerated such exploration. Such efforts are often targeted at producing a large series of structurally diverse molecules to be biologically tested through high throughput screening techniques in order to identify potential new drugs [1]. Despite the relatively reduced cost of parallel synthesis techniques, the selection of a relevant sample among all the possible substituents is required in order to adequately explore the experimental space using the minimum number of compounds [2]. The substituents introduced should produce the maximum coverage of the molecular properties,

avoiding the repetition of information. The primary aim of the mentioned sampling is to reduce the excessively large number of potential starting materials with the aim of generating a manageable synthetic plan [3]. For this purpose, we introduce in this article a new strategy of selection of compounds, which is based on their description by means of principal components obtained from variables including molecular weight, solvent-accessible surface area (SASA) and physicochemical descriptors estimated by quantum mechanical computations. Because of the simplicity of its computational process (PCA and cluster analysis) and the accessibility of the required software (a standard statistical package such as SPSS [4]), the present approach differs from other valuable approaches recently proposed, such as those based on self-organizing or Kohonen neural networks [5,6]. Our new approach has been tested in the selection of a sample of amines from those available in the Aldrich catalogue. This study constituted part of a real lead optimisation process in the framework of a drug development project.

The selection of representative members from a library of compounds in order to optimally cover the chemical diversity space is based on (usually computed) chemical properties [7]. With this purpose, many molecular descriptors have been proposed and thoroughly reviewed [8,9]. They can be classified in three groups:

- 1) Scalar molecular properties, which do not depend on a particular 3D conformation of the molecules, such as the molecular weight.

\*Address correspondence to this author at the Research Group on Biomedical Informatics, IMIM, Universitat Pompeu Fabra, C/ Dr. Aiguader 80, E-08003 Barcelona, Spain; Phone: +034 932 257 587; Fax: +034-932-213-237; E-mail: fsanz@IMIM.ES

<sup>†</sup>J.J. Lozano and H. G. de Terán contributed equally to this work.

<sup>#</sup>Current address: Dep. of Physiology and Biophysics, Mount Sinai School of Medicine, New York, USA.

- 2) Topological descriptors (“2D descriptors”), which rely on the molecular formulae, such as the connectivity indices.
- 3) Descriptors derived from 3D molecular structures, such as the dipolar moment.

In molecular diversity studies within the pharmaceutical field, it is usual to employ physicochemical descriptors [9], since they are often related to pharmacologically relevant molecular interaction phenomena (transport, ligand-receptor recognition, etc.) [10, 11].

As the information contained in each molecular descriptor is somehow correlated with the information contained in the other ones, principal components analysis (PCA) is a useful tool to eliminate redundant information [12], and has been often used to derive new orthogonal descriptors to be applied in molecular diversity studies [11,13-15]. The main principal components (PCs) allow a useful visualization of the diversity of a molecular library, and they can be used as input for maximum diversity sampling methods.

The methods for the selection of maximum diversity samples described in the literature use different computational techniques, such as hierarchical and non-hierarchical cluster analysis [16], genetic algorithms [17], neural networks [5,6], clique detection [18] or specifically designed methods like the most descriptive compound (MDC) method [19]. Random selection of the intended sample of compounds has been demonstrated to be less effective than the aforementioned rational algorithms [8,19]. Among them, in the present study we propose the use of non-hierarchical k-means cluster analysis [20], which will be compared with a distance-based algorithm [21].

## METHODS

### Molecular Library

The 923 primary amines included in the Aldrich catalogue were used in the present study. A molecular database containing their formulae (in sdf format) was available.

### 3D Structures Generation

The aforementioned molecular database containing the sdf planar coordinates of the 923 amines was first transformed into 923 independent files. The 923 formulae contained in the files were transformed into feasible 3D structures using the CORINA software [22], which is one of the fastest and more efficient software for such a purpose [23].

Taking into account that the selected amines will be used for generating substituents in a certain position of a molecular scaffold, the most meticulous approach would imply the substitution of one of the hydrogens of the amino group by the scaffold or a molecular fragment representing it. However, this operation could be neglected since the values of the molecular descriptors would be influenced in a similar

way by a molecular fragment shared by all the compounds, as showed previously by others [11]. This argument might be erroneous in the case of vectorial properties, such as the dipolar moment (DM) employed in the present study. To test the reliability of our structural simplification on this property, we recomputed the DM for nine amines homogeneously distributed along the dipolar moment range. In this second computation, we substituted one of the hydrogens of the amino group by a pyrazinyl moiety that represented the common scaffold of the parallel synthesis process that motivated the present analysis. The comparison of both dipolar moment values, those of the unsubstituted amines and those of the corresponding extended compounds, is presented in the Results section.

### Molecular Descriptors

The 923 3D molecular structures were used as input for quantum-mechanical computations, which were carried out at the AM1 level using AMSOL 6.3 software [24]. Molecules were fully optimised geometrically as solvated in water, using the SM5.4A solvation model [25]. Starting from the resulting geometries, single point computations were carried out to simulate the species solvated in octanol and benzene. Consequently, we compiled information about free energies of solvation in polar (water) and two non-polar (octanol and benzene) solvents.

The execution of quantum chemical computations produced strange results for 177 molecules because of one of the following reasons:

- a) They contained atoms such as Hg, Se, or B that make the amines uninteresting as drug-like fragments.
- b) The molecules were too large to be processed using quantum-mechanical (QM) methods.
- c) They were hydrochlorides, hydrobromides or other salts unsuitable for QM methods. On the other hand, the neutral forms of these compounds were also included in the database as different entries.
- d) In the case of enantiomers, which generate the same values of the physico-chemical descriptors considered, we selected only one of each pair.

The filtering of the above-described compounds could have been carried out by simple inspection of every compound included in the Aldrich database. Nevertheless, the present protocol allows for a certain automation of the process.

For the remaining 746 amines, six molecular descriptors were generated through the AMSOL computations. Two descriptors dealt with steric information (molecular weight and solvent-accessible surface area), one dealt with electronic information (dipolar moment), and three with solubility in solvents having different characteristics (free energies of solvation in water, benzene and octanol). The AMSOL program generates directly all these molecular descriptors

**Table 1. Bivariate Correlation Matrix Between the Original Descriptors**

	DM	MW	SASA	G_H <sub>2</sub> O	G_Bz
MW	0.18				
SASA	-0.15	0.75			
G_H <sub>2</sub> O	-0.44	-0.08	-0.01		
G_Bz	-0.19	-0.72	-0.76	-0.55	
G_Oct	-0.35	-0.46	-0.45	-0.87	-0.88

with no need of additional software, thereby simplifying the process.

Consequently, an initial data matrix having a 746 x 6 dimension was generated. As it is shown in the bivariate correlation matrix (see Table 1), all the molecular descriptors are correlated in different degrees. On the other hand, since they are expressed in different units, they cannot be mixed in certain computations (i.e., Euclidean distances).

### Principal Components Analysis

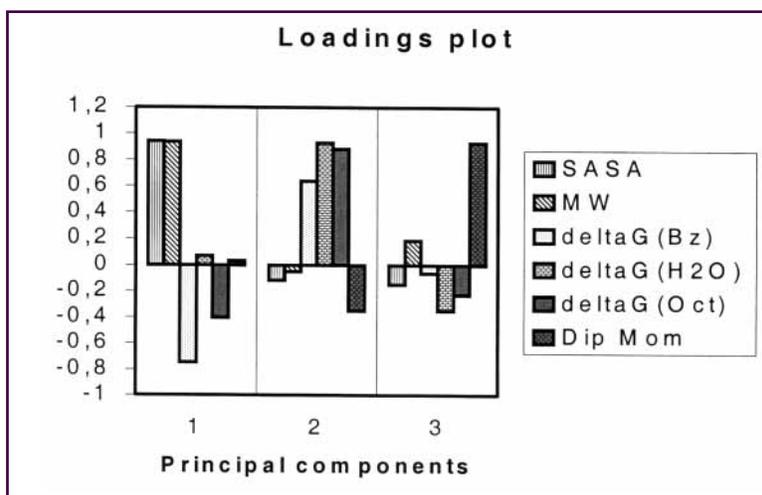
To solve the aforementioned problems of co-linearity between descriptors and the subsequent information redundancy, we propose the use of principal components (PCs) as molecular descriptors instead of the original ones [11,12-15]. The SPSS software [4] was used for this part of the analysis. The relevant number of PCs to be considered was determined on the basis of the percentage of the total variability of the data that had to be retained. The scores were standardized for the considered compounds in the PC space. A Varimax rotation of the selected PCs is advisable to obtain the maximum separation between the contributions of the original descriptors to each PC. The rotation aims to obtain high loadings for a subset of the original variables in each PC, while having the minimum overlap between such subsets (see Fig. 1). The orthogonality of the resulting PCs

implies a total lack of information redundancy in the new variables describing the compounds.

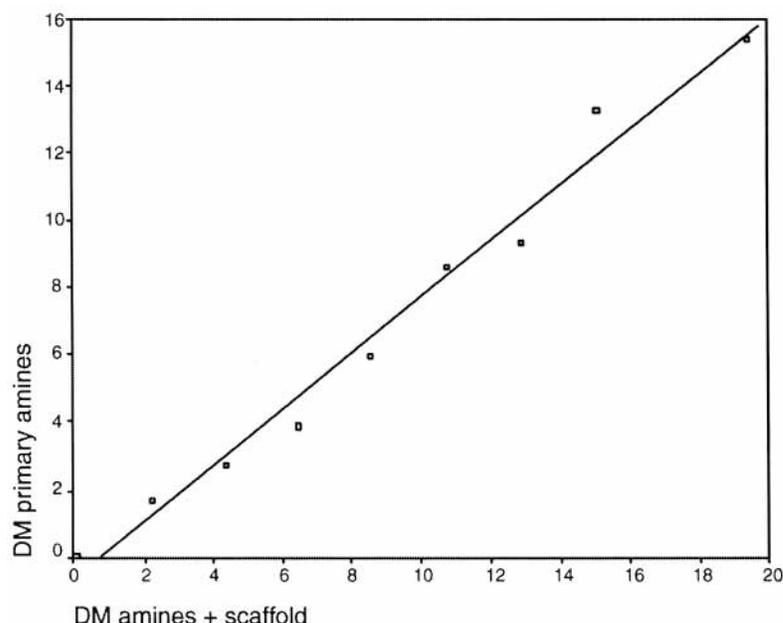
### Sampling Methods

The high diversity sample of  $n$  compounds was selected by a k-means cluster analysis [20] on the basis of the rotated PCs previously obtained. K-means is a non-hierarchical clustering method that starts with an arbitrary set of  $n$  centroids (vectors of descriptors), assigns the compounds to the closest centroid, recomputes each centroid as the mean of the vectors of descriptors of the corresponding compounds, and repeats iteratively the process until a stable classification is obtained. The compound closest to each final centroid is proposed as cluster representative. In the approach used in the present study, the initial set of centroids was randomly generated. The sample size (50 in this study) can be decided on the basis of statistical or feasibility criteria [16]. A critical aspect of k-means cluster analysis is its dependence of the initial set of centroids. We have studied the importance of this problem by comparing the results obtained from two different randomly generated initial sets of centroids.

In order to compare the proposed approach with another based on different algorithms, we used the longest minimum distance algorithm (LMD) developed by Marengo and Tordeschini [21]. The aim of this algorithm is iteratively



**Fig. (1).** PCA loadings plot. The contributions of each original molecular descriptor on each PC are depicted.



**Fig. (2).** Dipolar moment comparison: primary amines vs. secondary amines obtained by pyrazinyl substitution.

generating an  $n$ -sized sample of compounds having the largest possible distances between them. For this analysis, we employed the  $Q^2$  software [26] using standardized PCs without applying the Varimax rotation, which is not allowed in this software.

### Diversity Measures

In this kind of study, it is useful to assess the resulting solutions using diversity functions [27]. For such a purpose we used an  $R^2$ -like diversity coefficient defined as follows:

$$\text{Total diversity prior clustering} = D_T = \sum_{i=1}^N d_{ic}^2$$

$$\text{Diversity lost within the clusters} = D_{WC} = \sum_{j=1}^n \sum_{i=1}^{n_j} d_{ic_j}^2$$

$$R^2 = \frac{\text{Remaining diversity after clustering}}{\text{Total diversity}} = \frac{D_T - D_{WC}}{D_T}$$

being  $d_{ic}$  Euclidean distances computed in the PCs space,  $N$  and  $n$  the number of compounds of the whole set and the sample respectively,  $c$  the centroid of the whole set of compounds, and  $n_j$  and  $c_j$  the sizes and the centroids of the  $n$  clusters respectively. A refined, but computationally more difficult, version of the present coefficient would consist in defining  $c_j$  as the coordinates of the compound representing each cluster instead of its centroid. In the present study we have used the first definition.

## RESULTS AND DISCUSSION

### Structural Simplification

As explained in the Methods section, we evaluated the effect on the dipolar moment (MD) of the structural simplification consisting of not substituting one of the amino hydrogens by the relevant scaffold. The scatter plot of both MD values (simplified vs. extended structures) is presented in Fig. (2). The plot and the corresponding  $R^2$  value, which is equal to 0.980, show a good relationship between both series of values.

### Original Descriptors

Table 2 summarizes the values of the original descriptors in the entire dataset. Their distributions are clearly asymmetrical. Table 1 shows the correlations between such descriptors.

The descriptors look appropriate for drug discovery projects. For instance, free energies of solvation are effective in distinguishing molecules on the basis of their polarity [28], a feature that has well known importance in the pharmacological behaviour. Another parameter related to molecular polarity is the dipolar moment (DM). On the other hand, properties related to the molecular size, here represented by solvent-accessible surface area (SASA) and molecular weight (MW), are also known to affect biological behaviour. Thus, their ensemble may be useful as diversity descriptors in the analysis of series of drug candidates [29].

### Principal Component Analysis (PCA)

In the present case, considering the first three PCs retained 95% of the original variability of the compounds.

**Table 2. Descriptive Statistics of the Original Descriptors**

Descriptor	Maximum	Minimum	Mean	Median	SD
DM	19.39	0.08	4.48	3.82	3.04
MW	515.35	31.04	164.29	155.07	56.39
SASA	902.35	201.74	390.75	372.38	87.64
G(H <sub>2</sub> O)	0.44	-45.91	-10.11	-8.56	6.37
G(Bz)	-2.86	-26.09	-9.93	-9.43	3.32
G(Oct)	-4.50	-36.15	-12.63	-11.63	4.96

As can be seen in the loadings plot shown in Fig. (1), the orthogonal PCs resulting from the Varimax rotation had a clear interpretation:

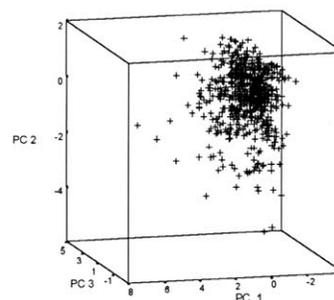
- PC1 was mainly constituted by steric information since the original variables that presented the loadings with the highest absolute values were solvent-accessible surface area (SASA) and molecular weight (MW).
- PC2 mainly included solvation information since the original variables that showed the coefficients with highest absolute values were the free energies ( G) of solvation in water, benzene and octanol.
- PC3 was mainly related to electrostatic information since the only original variable that resulted in a high loading in this component was the dipolar moment (DM).

The resulting descriptors (PCs) defined a standardized orthogonal three-dimensional space, which keeps 95% of the original information on the compounds.

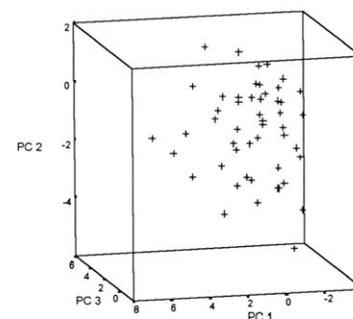
The distribution of the whole dataset in the space defined by the three PCs is visualized in the scores plot shown in Fig. (3a). It is worth noting that the density of compounds is greater in the intervals that correspond to low or moderated values of the PCs, contrasting with the low density of compounds having high values. This indicates that the initial set was not uniformly distributed in the molecular properties space.

### Selection of Compounds

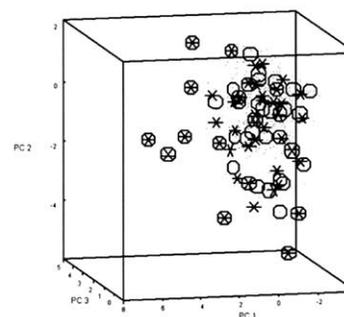
Starting from a set of 50 randomly generated centroids, the k-means cluster analysis produced a stable classification of the studied set of compounds into 50 groups. Obviously, the most populated clusters were those situated in the aforementioned high-density region, whereas the clusters created in the extreme zones were less populated or even single-compound groups. A high degree of chemical similarity between compounds belonging to the same cluster was observed by visual inspection. Even in the highly populated clusters, i.e., cluster 43 having 92 compounds-, a single representative of each cluster should be selected to



(a)



(b)



(c)

**Fig. (3).** Scatter plots in the PCs' space: (a) the 746 amines studied, (b) the 50 compounds selected by k-means cluster analysis, (c) superposition of the 50 compounds selected from two different runs of k-means cluster analysis (circles and asterisks; the entire initial population is represented by small points).

maximize the diversity of the final set [30]. The compound nearest to each final centroid was proposed as cluster representative. Fig. (3b) shows the position of each representative in the PCs space. Fig. (3c) depicts the comparison of the previous results with those obtained using an alternative initial set of centroids. Both solutions are coincident in the areas of the PCs' space with low density of compounds, whereas they differ in the particular compounds selected in the zones with high density of compounds. In

any case, the quality of both solutions in terms of  $R^2$ -like coefficient were the same (0.923).

To compare the solutions produced by the strategy above, another maximum diversity sample of 50 compounds was selected using an alternative distance based algorithm, the Longest Minimum Distance (LMD) method. This approach contains an iterative procedure analogous to the exchange algorithm used in D-optimal design, but using

**Table 3. Comparison Between k-Means and LMD Approaches**

k-means				LMD	
Cluster	Size	Compound	Dist. to centroid	Compound	Dist. to k-means centroid
1	40	HGT150	0.170	HGT111	0.676
2	11	HGT096	0.140	HGT093	0.328
3	1	HGT860	0.000	-	-
4	1	HGT331	0.000	HGT331	0.000
5	6	HGT784	0.192	HGT348	0.948
6	6	HGT079	0.245	HGT080	0.532
7	14	HGT873	0.121	HGT672	0.605
7				HGT517	0.717
8	36	HGT757	0.153	HGT889	0.515
9	3	HGT550	0.135	HGT921	0.655
10	1	HGT853	0.000	HGT853	0.000
11	47	HGT311	0.093	-	-
12	4	HGT338	0.105	HGT699	0.403
12				HGT410	0.568
13	7	HGT028	0.304	HGT028	0.304
14	12	HGT076	0.285	HGT091	0.578
15	9	HGT001	0.436	HGT842	0.798
15				HGT833	0.837
16	1	HGT545	0.000	-	-
17	11	HGT469	0.172	HGT469	0.172
18	14	HGT160	0.165	HGT662	1,112
19	25	HGT043	0.188	HGT278	0.593
20	30	HGT287	0.159	HGT217	0.674
20				HGT559	0.710
21	32	HGT465	0.190	HGT915	0.676
22	1	HGT074	0.000	HGT074	0.000
23	4	HGT846	0.209	HGT547	0.635
24	4	HGT845	0.224	HGT877	0.303
25	1	HGT220	0.000	HGT220	0.000
26	4	HGT528	0.279	HGT654	0.748

(Table 8). contd.....

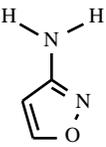
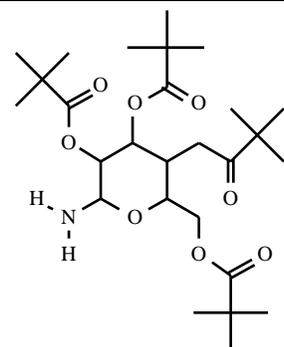
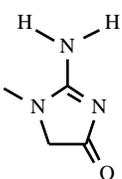
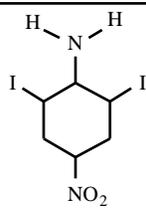
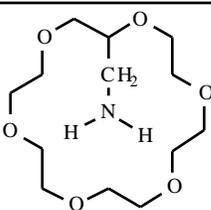
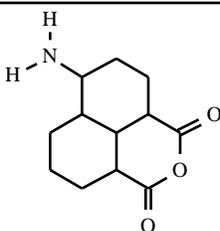
k-means				LMD	
Cluster	Size	Compound	Dist. to centroid	Compound	Dist. to k-means centroid
27	36	HGT693	0.114	HGT472	0.664
28	15	HGT823	0.171	HGT405	0.522
28				HGT356	0.540
29	24	HGT374	0.107	HGT389	0.501
30	22	HGT265	0.062	-	-
31	2	HGT755	0.474	HGT755	0.474
32	1	HGT066	0.000	HGT066	0.000
33	23	HGT121	0.318	-	-
34	17	HGT772	0.209	HGT535	0.513
34				HGT189	0.901
35	1	HGT409	0.000	HGT409	0.000
36	11	HGT330	0.370	HGT109	0.418
37	8	HGT092	0.291	HGT077	0.582
38	19	HGT378	0.173	HGT394	0.942
39	1	HGT522	0.000	HGT522	0.000
40	2	HGT897	0.340	-	-
41	1	HGT881	0.000	HGT881	0.000
42	4	HGT529	0.231	HGT027	0.930
43	92	HGT851	0.074	HGT495	0.681
44	42	HGT413	0.098	-	-
45	4	HGT171	0.399	HGT171	0.399
46	10	HGT308	0.169	HGT309	0.460
46				HGT557	0.917
47	13	HGT118	0.147	HGT319	0.706
48	2	HGT700	0.227	HGT035	0.227
49	31	HGT370	0.210	HGT864	0.633
50	40	HGT022	0.100	HGT796	0.638

distance considerations with the purpose of obtaining designs well distributed in the variable space [21]. The correspondence between the representatives obtained here and the groups generated with the k-means algorithm is shown in Table 3. This table shows the size of each cluster, the compounds selected by k-means and LMD that belong to the same cluster, as well as the distances to the corresponding centroids. In seven cases, LMD proposed two compounds belonging to the same k-means cluster. Vertical bold lines in Table 3 indicate them. Consequently, since the sample size was the same in both approaches, there were seven k-means clusters without LMD representative. In general, the clusters that contained two compounds proposed by LMD were the most populated. In such cases, it has to be pointed out that

both representatives showed long Euclidean distances to their centroid and relatively short distances to that of the nearest cluster non-represented by LMD. An 86% of agreement between both methods was found. Agreement was defined as the percent of clusters generated by k-means analysis that included at least one compound generated by the alternative LMD algorithm. Finally, it has to be pointed out that the solution obtained with the LMD method had the same  $R^2$ -like diversity coefficient (0.923) as the results obtained with the strategy postulated in the present study.

As an example of the compounds selected as cluster representatives, the formulae of those having extreme values of each PC are shown in Table 4. It can be appreciated that

**Table 4. Cluster Representatives with Extreme Values on the Principal Components**

PC	Lowest value		Highest value	
	Cluster	Compound	Cluster	Compound
PC1 (steric)	48		10	
PC2 (solvation)	31		25	
PC3 (dip. moment)	41		39	

the proposed compounds have completely different values of the molecular properties represented by each PC. Thus, the first line of the table shows two compounds having very diverse steric characteristics, while the last line shows two compounds with very different charge balances. The two compounds with extreme values of the second PC, despite being diverse, are more difficult to interpret, which is logical taking into account that this PC is a mix of solvation thermodynamics in solvents exhibiting different polarities.

## CONCLUSIONS

We have developed and tested a fast, semiautomatic distance-based strategy for the selection of a maximum diversity sample of compounds from a relatively large initial dataset. This approach uses chemical information that is relevant for the biological behaviour and avoids information heterogeneity and redundancy by using standardized Varimax-rotated principal components instead of the original molecular descriptors.

## ACKNOWLEDGMENT

The authors thank Dr. M. Pastor for helpful discussions. CIRIT is also acknowledged for the grant provided to one of the authors.

## LIST OF ABBREVIATIONS

- PCA = Principal components analysis  
 MW = Molecular weight  
 SASA = Solvent-accessible surface area  
 DM = Dipolar moment  
 $G(\text{H}_2\text{O})$  = Free energy of solvation in water  
 $G(\text{Bz})$  = Free energy of solvation in benzene  
 $G(\text{Oct})$  = Free energy of solvation in octanol  
 LMD = Longest minimum distance

## REFERENCES

- [1] Special issue on "Computational methods for the analysis of molecular diversity", *Perspect. Drug Discov. Design*, **1997**, 7/8.
- [2] Newton, C.G. In *Molecular diversity in drug design*; Dean, P. M.; Lewis, R. A. Eds.; Kluwer Academic Publishers: Dordrecht, **1999**; Ch 2, pp. 23-42.

- [3] Gund, P.; Sigal, N.H. *Immunol. Today*, **1999**, 2(Suppl. 1), 25.
- [4] SPSS 9.0, SPSS Inc. Chicago.
- [5] Polanski, J.; Jarzembek, K.; Gasteiger, J. *Comb. Chem. High Throughput Screen.*, **2000**, 3, 481.
- [6] Ros, F.; Audouze, K.; Pintore, M.; Chretien, J.R., *SAR QSAR Environ. Res.*, **2000**, 11, 281.
- [7] Spellmeyer, D.C.; Grootenhuis, P.D.J. *Ann. Rep. Med. Chem.*, **1999**, 34, 287.
- [8] Gillet, V.J. In *Molecular diversity in drug design*; Dean P. M.; Lewis, R. A. Eds.; Kluwer Academic Publishers: Dordrecht, **1999**; Ch 3, pp. 43-65. And references there in.
- [9] Downs, G. M.; Willett, P. *J. Chem. Inf. Comput. Sci.*, **1994**, 34, 1094.
- [10] Kubinyi, H. *QSAR: Hansch analysis and related approaches*; VCH: Weinheim, **1993**.
- [11] Gibson, S.; McGuire, R.; Rees, D.C. *J. Med. Chem.*, **1996**, 39, 4065.
- [12] Joliffe, I.T.; Morgan, B.J.T. *Stat. Methods Med. Res.*, **1992**, 1, 69.
- [13] Langer, T.; Hoffman, R.D. *Quant. Struct. -Act. Relat.*, **1998**, 17, 211.
- [14] Eriksson, L.; Johansson, E. *Chemom. Intell. Lab. Sys.*, **1996**, 34, 1.
- [15] Oprea, T.; Gottfries, T., ChemGPS, a chemical space navigation tool. 13<sup>th</sup> European Symposium on QSAR; Düsseldorf, **2000**.
- [16] Voght, W.; Nagen D.; Sator, H. *Cluster analysis in clinical chemistry: a model*; J Wiley & Sons, Ltd.: Chichester, **1987**.
- [17] Sheridan, R.P., SanFeliciano, S.G.; Kearsley, S.K.; *J. Mol. Graph.*, **2000**, 18, 320.
- [18] Gardiner, E. J.; Holliday, J.D.; Willet, P.; Wilton, D.J.; Artymiuk, P.J. *Quant. Struct.-Act. Relat.*, **1998**, 17, 232.
- [19] Hudson, B. D.; Hyde, R.M.; Rahr, E.; Wood, J. *Quant. Struct.-Act. Relat.*, **1996**, 15, 285.
- [20] Späth, H. *Cluster-Analyse-Algorithm zur Objektklassifizierung und Datenreduction*, 2<sup>nd</sup> revised Ed., R. Oldenbourg Verlag: Munich, **1977**.
- [21] Marengo, E.; Todeschini, R. *Chemom. Intell. Lab. Sys.*, **1992**, 16, 37.
- [22] Gasteiger, J.; Rudolph, C.; Sadowski, J. *Tetrahedron Comp. Method.*, **1990**, 3, 537.
- [23] Sadowski, J.; Gasteiger, J.; Klebe, G., *J. Chem. Inf. Comput. Sci.*, **1994**, 34, 1000.
- [24] Hawkins, G.D.; Giesen, D.J.; Lynch, G.C.; Chambers, C.C.; Rossi, I.; Storer, J.W.; Li, J.; Rinaldi, D.; Liotard, D.A.; Cramer C.J.; Truhlar D. G., AMSOL-version 6.5.2, University of Minnesota, Minneapolis, **1997**.
- [25] Chambers, C.C.; Hawkins, G.D.; Cramer, C.J.; Truhlar, D. G. *J. Org. Chem.*, **1996**, 61, 8720.
- [26] Q2 4.5. Multivariate Infometric Analysis srl. Perugia, Italy, **1999**.
- [27] Hassan, M.; Bielawski, J.P.; Hempel, J.C.; Waldman, M. *Mol. Divers.*, **1996**, 2, 64.
- [28] Cummins, D. J.; Andrews, C.W.; Bentley, J.A.; Cory, M. *J. Chem. Inf. Comput. Sci.*, **1996**, 36, 750.
- [29] Patterson, D. E.; Cramer, D.E.; Ferguson, A.M.; Clark, R.D.; Weinberger, L.E. *J. Med. Chem.*, **1996**, 39, 3049.
- [30] Bailey, D.S.; Furness, L.M.; Dean, P.M. *Immunol. Today*, **1999**, 2(Suppl. 1), 6.



## **ARTÍCULO II**

Artículo II:

NOVEL APPROACHES FOR MODELING OF THE A1  
ADENOSINE RECEPTOR AND ITS AGONIST BINDING SITE

HugoGutiérrez de Terán, Nuria B. Centeno, Manuel Pastor, FerranSanz

Consultable en:

Gutierrez-de-Teran H, Centeno NB, Pastor M, Sanz F. "Novel approaches for modeling of the A1 adenosine receptor and its agonist binding site" *Proteins*. 2004 Mar 1;54(4):705-15



## **ARTÍCULO III**



# Comparative analysis of putative agonist binding modes in the human A<sub>1</sub> adenosine receptor

Hugo Gutiérrez-de-Terán<sup>[a]</sup>, Manuel Pastor<sup>[a]</sup>, Nuria B. Centeno<sup>[a]</sup>, Johan Åqvist<sup>[b]</sup> and Ferran Sanz<sup>[a]</sup>\*

Key words: adenosine receptors; binding free energy; docking exploration; molecular dynamics; homology modeling.

**Chembiochem, accepted for publication 29/12/2003**

---

<sup>[a]</sup> Prof. F. Sanz, Dr. M. Pastor, Dr. N.B. Centeno, H. Gutiérrez-de-Terán  
Research Group on Biomedical Informatics (GRIB)  
Institut Municipal d'Investigació Mèdica, Universitat Pompeu Fabra  
Carrer Dr. Aiguader 80,  
E-08003 Barcelona, Spain  
Fax: (+34)-93-224 0875  
e-mail: [fsanz@imim.es](mailto:fsanz@imim.es)

<sup>[b]</sup> Prof. J. Åqvist  
Department of Cell and Molecular Biology  
Uppsala University  
BMC, Box 596 SE-751 24  
Uppsala, Sweden  
Fax: +46-(0)18 - 53 69 71  
e-mail: [aqvist@xray.bmc.uu.se](mailto:aqvist@xray.bmc.uu.se)

## ABSTRACT

In a recent study, a model of the human A<sub>1</sub> adenosine receptor and its agonist binding site was reported, where two putative binding modes in the same binding site were proposed for the natural agonist, adenosine. The present work investigates the flexibility of this binding site by exhaustive exploration with the natural agonist and three other adenosine derivatives, N<sup>6</sup>-cyclopentyladenosine (CPA), 2-chloro-N<sup>6</sup>-cyclopentyladenosine (CCPA) and 5'-N-ethylcarboxamidoadenosine (NECA). Our aim was to find a common binding mode for agonists that explains the role in the binding process of the different substitutions allowed at the 2', N<sup>6</sup> and 5' positions of adenosine. This problem was addressed through docking simulations, molecular dynamics studies and estimations of the ligand binding free energy both with AUTODOCK scoring function and the linear interaction energy (LIE) approach. The results point towards a single receptor binding position that explains the effects of the different chemical modifications on the adenosine derivatives considered here.

## INTRODUCTION

The nucleoside adenosine exerts its extracellular effects through coupling with transmembrane adenosine receptors (ARs), a family of the G protein coupled receptors (GPCRs). In the last decades much effort has been made pursuing pharmacological understanding of the four adenosine receptor subtypes, named A<sub>1</sub>, A<sub>2a</sub>, A<sub>2b</sub> and A<sub>3</sub>. Among them, the A<sub>1</sub> subtype is of particular interest, both because it is widely studied<sup>[1]</sup> and because of the physiological functions in which it is involved, making it attractive as a potential drug target<sup>[2,3]</sup>. In this respect, A<sub>1</sub>AR agonists can be of therapeutical interest in different biological systems: in the CNS, this receptor is located on cortex, hippocampus, cerebellum and thalamus, where its activation leads to neuroprotective effects through the blockade of neurotransmitter release.<sup>[4]</sup> In the heart, A<sub>1</sub>AR agonists cause negative chronotropic, dromotropic and inotropic effects, which makes them attractive as potential cardioprotective and anti-infarct agents.<sup>[5]</sup>

Selective and potent agonists have been developed for the A<sub>1</sub> receptor, providing tools to a better understanding of its pharmacology. Agonists for ARs are chemically related with the natural agonist adenosine (see Table 1). The ribose moiety appears to be essential for both affinity<sup>[2]</sup> and agonistic profile<sup>[6]</sup> of adenosine derivatives. Single modifications on the either 2' or 3' hydroxyl group lead to molecules that show a profile of partial agonist,<sup>[7,8]</sup> while removal of both substituents at the same time confers antagonistic properties.<sup>[9]</sup> Modifications of the ribose moiety maintaining a full agonist profile are limited to 5'-uronamido derivatives, from which NECA (5'-*N*-ethylcarboxamidoadenosine) is a potent non A<sub>1</sub> selective example.<sup>[10]</sup> The introduction of small aromatic or cycloalkyl substituents at the N6 position leads to more potent and selective A<sub>1</sub> derivatives, such as N6-cyclopentyladenosine (CPA).<sup>[2]</sup> Another position of substitution is the 2-position, where a chlorine atom combined with N6 substitution yielded CCPA (2-chloro-N6-cyclopentyladenosine),<sup>[11]</sup> the most potent agonist known for the A<sub>1</sub>AR.<sup>[12]</sup> Thus, molecules such as NECA, CPA and CCPA have been used for years to study this receptor subtype (see references [6,13] and references therein).

In order to investigate the location of the agonist binding site and the role of several aminoacids known to influence ligand binding, we have recently developed a model for the human A<sub>1</sub> adenosine receptor<sup>[14]</sup> (hA<sub>1</sub>AR) based on the X-ray crystal structure of bovin rhodopsin.<sup>[15]</sup> Using this model and the GROUP module of the GRID software<sup>[16]</sup>, we have exhaustively explored the putative ribose binding site. A single solution was found, located between helices 1, 2 and

7 and composed by residues Ser1.46, Asp2.50, His7.43 and Ser7.46. The high polarity of this pocket is able to establish enough hydrogen bond (H-bond) interactions to balance the ribose desolvation energy, and also showed high similarity with other experimentally determined ribose binding sites. This ribose binding pocket was a starting anchoring point for the search of the binding position for the natural agonist, adenosine (ADO), carried out by docking and molecular dynamics (MD) simulations, concluding that two adenosine binding modes were possible, both involving helices 1, 2, 3, 6 and 7 and almost the same residues. One binding mode presented the adenosine amino group facing the extracellular side, interacting with Thr7.42, while the other complex obtained linked the amino group with another threonine (Thr3.36) located in helix 3, therefore looking towards the central core of the helical bundle. The adenine moiety was in both cases stabilized by hydrophobic interactions with Val3.32, a conserved residue in ARs that corresponds to the aspartate that binds the basic nitrogen of the ligands in aminergic receptors.<sup>[17]</sup> Both binding modes may explain mutagenesis data concerning residues such as Thr3.36,<sup>[18]</sup> or His7.43,<sup>[19,20]</sup> and consequently no criteria were found in the precedent study for determining the preferred binding mode.

The present work performs a deeper and comparative analysis of such binding modes by considering the docking of other relevant A<sub>1</sub> agonists. This analysis is intended to explore structure-activity relationships of adenosine derivatives, identifying the role of the different chemical substitutions and their relative counterparts in the receptor. Moreover, the obtained docking model should be predictive for the differences in potency of the ligands studied. To make progress towards this goal, we made use of different computational simulation methods like docking, molecular dynamics and binding free energy calculations.

## **METHODS**

### **Human A<sub>1</sub> adenosine receptor model**

In the present calculations, the human A<sub>1</sub> adenosine receptor model (hA<sub>1</sub>AR) previously reported<sup>[14]</sup> was used. This model incorporates only the seven helix bundle, discarding the loop regions. Briefly, each helix was separately built and geometrically optimized, followed by the packing of the seven helices using the experimental structure of bovin rhodopsin<sup>[15]</sup> as a template. Finally, the model was energy optimized and refined using molecular dynamics techniques. The generalized numbering scheme proposed by Ballesteros and Weinstein<sup>[21]</sup> was used to name the residues through the text. This nomenclature allows an easier identification of the residues within the seven-

helix bundle, as well as a simpler association of equivalent residues of different GPCRs or of the same receptor in different species.

### **Modeling of the ligands**

The four agonists studied, ADO, NECA, CPA and CCPA are depicted in Table 1, together with their experimentally determined binding data. The 3D structures of adenosine derivatives were built by modifying the crystallographic coordinates of adenosine (refcode ADENOS10 in the Cambridge structural database<sup>[22]</sup>) with appropriate substituents using the Builder module of InsightII.<sup>[23]</sup> Each molecule was then energy minimized with the semi empirical AM1 Hamiltonian as implemented in MOPAC93.<sup>[24]</sup> Final geometries and partial charges obtained in this step were used as input for AUTODOCK computations.

In order to find necessary force field parameters for the molecular dynamics and binding free energy calculations, we performed single point ab-initio calculations with the GAUSSIAN98 suite of programs<sup>[25]</sup> using the Hartree-Fock method and the 6-31G\* basis set. The restrained electrostatic potential fitting procedure<sup>[26]</sup> was applied on the calculated electrostatic potentials in order to obtain partial atomic charges compatible with the Amber Parm94<sup>[27]</sup> force-field.

### **Docking studies**

A docking exploration with AUTODOCK3.0<sup>[28]</sup> was performed for each considered agonist. This program allows full flexibility for the ligands, while keeping the geometry of the receptor frozen. The exploration of docking positions was carried out through 100 runs of the Lamarckian genetic algorithm (LGA), using AUTODOCK default parameters. The resulting docking positions were clustered according to an RMS criterion of 1 Å. The force field implemented in the program was modified by using a united-atom model for both protein and ligand, and 12-10 hydrogen-bond parameters for all the O···H and N···H interactions, whenever the involved nitrogen atoms have a lone pair. We restrained the docking exploration to the region containing the residues experimentally known to influence agonist binding, by defining a 20<sup>3</sup> Å<sup>3</sup> grid centered on Nδ of His7.43.

### **Molecular dynamics and Binding free energy predictions**

In addition to the empirical binding energies provided by AUTODOCK, free energies of binding were calculated using the linear interaction energy (LIE) method.<sup>[29,30]</sup> This method is based on thermal conformational sampling of the ligand, both in the free state (i.e., solvated in water) and bound to the solvated protein. The estimated energy of binding is calculated as a linear combination of the

differences in the average ligand-surrounding interactions (where surrounding is referred to protein and water, when considering the bound state, or just water if considering the free state). Interaction energies are split into electrostatic and van der Waals terms, and weighted by different factors:

$$\Delta G = \alpha \left( \langle V_{l-s}^{vdW} \rangle_p - \langle V_{l-s}^{vdW} \rangle_w \right) + \beta \left( \langle V_{l-s}^{el} \rangle_p - \langle V_{l-s}^{el} \rangle_w \right) + \gamma \quad (1)$$

Where the  $\langle \rangle$  terms denote thermal averages sampled during molecular dynamics of the electrostatic (el) and van der Waals (vdW) energies for the ligand atoms in the protein (p) and water (w) environments. The coefficients in the above equation are taken as  $\beta = 0.5$ , when the linear electrostatic response approximation is valid,<sup>[31]</sup> and  $\alpha = 0.181$  for the non-polar interactions. In our case, given that all ligands are neutral molecules with two or more hydroxyl groups (which has been shown to cause deviations from exact linear response)<sup>[32]</sup> the parameter  $\beta$  was set to the earlier determined value of 0.33.<sup>[30]</sup> The constant parameter  $\gamma \neq 0$  can be adjusted by least squares optimization in order to quantitatively improve the calculated absolute binding energies with respect to the experimental ones. Such a constant term appears to be dependent on the nature of binding site studied and this issue is discussed elsewhere.<sup>[33]</sup> Here, we are mainly concerned with relative binding energies, and also have only a few compounds to work with, so that a detailed optimization of this parameter is not warranted.

The program Q<sup>[34]</sup> was used for the MD simulations and analysis, employing the Amber parm94 force field<sup>[27]</sup> implemented in the program. In both the "bound" and "free" states, the ligand was solvated with a randomly oriented TIP3P<sup>[35]</sup> water sphere of 18 Å radius, in which water molecules were confined to the sphere through a radial restraining potential.<sup>[34,36,37]</sup> In the "bound" simulation, ionic groups of the protein within this sphere that are not too close to the boundary (Asp2.50 and Glu1.39) were modeled as charged. The resulting net charge for the sphere of simulation was -2 since there were no Arg or Lys residues that could be ionized in order to counter-balance negative charges. Furthermore, neither of the histidines inside the simulation sphere (His7.43 and His6.51) are likely to be positively charged according to experimental data.<sup>[38]</sup>

The solvent was relaxed using a strong thermal bath coupling ( $\tau = 1$  fs), a temperature of 0 K, and restraining solute atoms with a force constant that decreased from 50 to 25, 10 and 5 kcal mol<sup>-1</sup>Å<sup>-2</sup> in blocks of 0.8 ps each. This step, which is very similar to a steepest descendent energy minimization, was followed by a heating phase consisting of blocks of 6 ps at 50 K, 150 K and 300 K, while relaxing the bath coupling ( $\tau = 5$  fs) and maintaining the 5 kcal mol<sup>-1</sup>Å<sup>-2</sup> force constant on all the protein and ligand atoms. In some cases, a set of five simulated annealing (SA) cycles was performed at this point, consisting on 4 ps smooth ( $\tau = 50$  fs) warming up until 900 K followed by a fast cooling down similar to energy minimization ( $\tau = 5$  fs), in order to allow the system to explore conformational changes that required crossing high-energy barriers from the initial docking position. A harmonic constraint of 10 kcal mol<sup>-1</sup>Å<sup>-2</sup> was applied on the C $\alpha$  protein atoms in this step to keep the trace of the protein fixed while relaxing the ligand and protein sidechains.

Productive MD simulations for the data collection were performed at 300 K (coupling to the temperature bath was set to  $\tau = 10$  fs) with a time step of 1.5 fs. No cut-off was applied to the ligand interactions, while a 10 Å cut-off was applied to other non-bonded interactions, together with the local reaction field approach<sup>[39]</sup> for long-range electrostatic interactions beyond that cut-off. A 10 kcal mol<sup>-1</sup>Å<sup>-2</sup> force constant was imposed to the atoms in the boundary zone between 16.5 and 18 Å from sphere center and, due to the absence of loops and the membrane, a force constant of 5 kcal mol<sup>-1</sup>Å<sup>-2</sup> was maintained for the alpha carbons inside the sphere. (Note, that due to the size of the protein no water penetrates into the membrane region) The rest of the protein inside that sphere and all ligand atoms were completely free to move, while the protein atoms outside the sphere of simulation were fixed by a high (100 kcal mol<sup>-1</sup>Å<sup>-2</sup>) harmonic constraint. Energy data were collected at this stage for 300-1500 ps starting from a moment at which the system could be considered equilibrated. The length of the sampling stage was such than the energy data could be split in two halves with equal average energies.

In the "free" simulation, the solvent was first relaxed by means of 0.2 ps MD using a strong bath coupling ( $\tau = 0.2$  fs) followed by 15 ps (relaxing bath coupling to  $\tau = 10$  fs), at temperature of 300 K and applying a harmonic constraint (10 kcal mol<sup>-1</sup>Å<sup>-2</sup>) on all ligand atoms. For the data collection, the central atom of the ligand was geometrically restrained to the center of the grid through

a 50 kcal mol<sup>-1</sup>Å<sup>-2</sup> force constant, to ensure a homogenous solvation. Time step and bath coupling were set to the same values than in the "bound" state.

Analysis of the MD results was performed with the tools provided in the Q software plus additional visual inspection with VMD.<sup>[40]</sup>

## RESULTS

The starting point for the present study was the previously reported hA<sub>1</sub>AR adenosine binding site that, given the relative symmetry of H-bond acceptor groups in adenosine derivatives,<sup>[41]</sup> was able to accommodate two potential agonist binding modes.<sup>[14]</sup> In order to deal with this ambiguity, we picked three classic and potent full A<sub>1</sub>AR agonists for which the experimental affinity constant had been determined for the human receptor (NECA, CPA and CCPA), as well as the natural ligand, adenosine (see Table 1), and studied their binding to our hA<sub>1</sub>AR model. This was done in a two step protocol as follows: first, an automatic docking exploration provided ideas about the most relevant binding positions for each compound. Then, each relevant position was further analyzed by MD and an estimation of the free energy of binding was performed using the linear interaction energy (LIE) approach.

### Docking studies

An exhaustive docking exploration of the binding site with each of the four ligands considered was performed with the AUTODOCK software.<sup>[28]</sup> This program has been previously used for docking flexible ligands into models of GPCRs,<sup>[42-44]</sup> and its usefulness has been further tested by us in the experimental case of bovine rhodopsin.<sup>[14]</sup>

Results of docking explorations are summarized in Table 2. For each of the ribose unmodified ligands (ADO, CPA and CCPA), two binding positions appear in agreement with the results reported in our previous article. Both positions show the adenosine scaffold located in the cavity between helices 2, 3 and 7 and involve essentially the same residues, but playing different roles in agonist binding in each case. These docking positions will hereafter be referred to as "position A" and "position B". (see Fig 1) It should be pointed out that other docking positions were also found for each ligand, but none of them showed consistency among the ligand series and/or agreement with mutagenesis data, comprising low-quality solutions not selected therefore for further investigation. In position A, docking positions of ADO and CPA show the 5'-hydroxyl of adenosine interacting with the carboxylic acid of Asp2.50, whereas the OH3' and OH2' groups are surrounded by Ser1.46 His7.43 and Ser7.46, and the N6 amino group points towards

Thr3.36. In this orientation, the hydrophobic pocket that accommodates the N6 hydrophobic substitution (N6-cyclopentyl in the case of CPA and CCPA) is composed by Leu3.33, Thr3.36, Leu6.51 and Thr7.42 (implying the methyl groups sidechain in the case of threonines). The CCPA ligand is predicted to dock slightly closer to the extracellular part of the receptor, thus missing interactions with helices 1 and 2. The substitution at position 2 (chlorine in CCPA) is pointing towards the extracellular part of the receptor in the region between helices 1 and 7. According to the AUTODOCK scoring function, this docking solution is isoenergetic for the three ribose unmodified ligands, thus not explaining the experimental increase in affinity found when 2 and N6 substitutions are present. Furthermore, this binding mode was not found for NECA, probably due to a steric hindrance between the 5' substituent and Asp2.50. Regarding the other orientation (position B), AUTODOCK provided this solution as the best ranked binding position for all the ligands except for ADO (for which the scoring energies for positions A and B are very similar though). In this case, Asp2.50 interacts with the ribose group through the 2' and 3' hydroxyl groups, also surrounded by Ser1.46 and Ser7.46 and the 5' hydroxyl points towards the backbone oxygen of His7.43. The amino group is close to residue Thr7.42, thus pointing up in this orientation to the extracellular side of the receptor. The hydrophobic pocket needed for the N6 substitution is constituted by Leu6.51, Ile7.39, Thr7.42 and to a lesser extent by Leu3.33 and Thr7.35. These hydrophobic interactions, related to the structural difference between CPA and ADO, imply an increase in the free energy of binding, both experimentally and estimated by the AUTODOCK scoring function (Table 2). The 2-Cl substituent of CCPA is accommodated between residues Thr3.36 and Trp6.48, looking in this case towards the cytosolic side of the receptor. However, this chemical modification is not translated into an increase of the estimated receptor affinity with respect to CPA as expected from experimental data, and on the contrary, a small decrease is predicted. Regarding the O5' position of adenines, when a carboxamido group is present (5'-N-ethylcarboxamide in NECA), the NH points towards His7.43 whilst the ethyl chain is located between helices 1 and 2. The estimated free energy of binding of NECA lies between the values calculated for ADO and CPA, in accordance with experimental data.

These docking results point to position B as the one that best explains the experimental results, because of several reasons: first, the ranking within the series obtained in this position with the scoring function estimation of the free energy of binding is fairly close to the experimental

one. Additionally, in all cases the estimated affinity is greater in position B than in position A (with the exception of equal energetic values for ADO). Finally, position A was not automatically found for NECA and the ligand CCPA did not fit well with respect to both ADO and CPA. The correct ranking of compounds in the docking position B is disrupted by the pair CPA-CCPA, the rank of which is inverted. These two ligands present the greatest similarity among the dataset, both in structure and in experimental  $\Delta G$  ( $\Delta\Delta G = 0.6 \text{ kcal mol}^{-1}$ ), and the prediction of such a small difference may be out of the scope of the AUTODOCK scoring function, especially when the only structural difference lies on a chlorine atom, which could be deficiently parametrized.

### **Molecular dynamics and binding free energy predictions**

When attempting to predict binding positions for a certain ligand into a protein, an automatic docking exploration provides good starting points for further detailed analyses of the stability of ligand-protein complexes and the interactions involved in binding. Such analyses are preferably done by molecular dynamics (MD) simulations, thus allowing to take into account the effects of the receptor flexibility and the explicit water solvation on the complex of interest. With the LIE approach, it is possible to estimate ligand binding free energies from MD simulation averages of the non-bonded intermolecular ligand potential energies. In our case, these estimated binding free energies might not be expected to quantitatively reproduce the absolute values for each complex for several reasons. First, it has been shown earlier that hydrophobic binding sites require a negative value of the constant  $\gamma$  in Equation (1), which in the case of a very hydrophobic site such as that of retinol binding protein can reach a value of about  $-7 \text{ kcal mol}^{-1}$ .<sup>[33]</sup> Second, the fact that we are working here with a homology model that does not incorporate loop regions and the membrane environment of the hA<sub>1</sub>AR structure may have some effect on the absolute  $\Delta G$  scale of the calculated energies. Nevertheless, the influence of these features is expected to be the same through the whole dataset, since the ligands considered are full agonists and consequently they have to share a common binding site. Thus, the differences in binding free energy and the ranking predicted for the series of agonists should in our opinion be reliable, as has been found in previous studies,<sup>[33,45]</sup> and therefore be informative about the relevance of each binding mode proposed. In this context, it has to be noted that the LIE method without reparametrization has been successfully used to predict relative binding free energies within the  $1 \text{ kcal mol}^{-1}$  range.<sup>[33,45]</sup> However, methods using conformational sampling

are always associated with convergence errors and it is essential to provide estimates of these in order to judge whether predicted binding energy differences are statistically significant.

The free energy of binding estimation obtained with the AUTODOCK scoring function for the first docking position studied (position A) was not very encouraging (see Table 2). In fact, the influence of the different chemical substitutions in the adenosine scaffold is not reflected in any improvement of the affinity of the compounds. Due to the absence of a good superposition of the four agonists after AUTODOCK computations (that locate CCPA slightly closer to the extracellular side than the other ligands and that does not predicted binding for NECA) the MD simulations of CCPA and NECA started from the geometry found in the CPA and ADO simulations, after mutating the corresponding ligand structures.

The LIE predictions of ligand affinities do not predict the ranking of binding free energies better than AUTODOCK in this binding position A (see Table 3). In fact, only the LIE estimation of CCPA binding free energy showed this compound with a slight improvement with respect to the natural agonist (ADO), while both CPA and NECA were predicted to bind worse than ADO (positive values of  $\Delta\Delta G$ , in disagreement with experimental data). A careful analysis of the protein-ligand interactions reveals a common pattern of interactions between the ribose and the protein. The hydrogen bond requirements of the ribose moiety in the protein binding site are not well satisfied and only a single strong interaction was achieved between OH5' and Asp2.50. The other two hydroxyl groups (at 2' and 3' positions) display H-bonds only with low occurrence with surrounding sidechains (His7.43 and Ser1.40 and Asp2.50 respectively), the first hydroxyl being mostly involved in intramolecular interactions. This binding profile is different in the ribose modified ligand NECA. The MD simulation of this last ligand started from the corresponding mutation of adenosine to NECA, in such a way that the 5'-N-ethylcarboxamido substitution was located between helix 2 (Asp2.50), 3 (Leu3.35) and 6 (Phe6.44). The 2' hydroxyl group of NECA shows a stable H-bond with His7.43, and the adenine core is well overlapped with that of the ribose unmodified ligands. However, these new contacts are not able to explain the experimental increase of affinity due to the 5'-N-alkylcarboxamido modification of adenosine. Taking into account the high polarity of the ribose, it is expected that the ligand-receptor complex should be able to generate enough H-bond interactions to balance the ribose desolvation energy. In contrast, the electrostatic energy term is more favorable (more negative

values for  $\langle V_{l-s}^{el} \rangle$ ) in water than in the binding position A for the four agonists studied (Table 3). Additionally, all the ligands adopt a south conformation (Fig. 2) in this binding position, with an intramolecular hydrogen bond between OH2'...N3. This feature is in disagreement with the results of conformational analysis of ribose-modified adenosine derivatives,<sup>[46]</sup> which show a preference for the north-*anti* conformation. The adenine core in all the four cases makes favorable hydrophobic interactions with Val3.32 and an H-bond between N3 and His7.43 is intermittently achieved. The behavior of the N6 position is stable along the MD simulation, with a single H-bond with Thr3.36 and, in the case of CPA and CCPA, favorable hydrophobic interactions with the previously described hydrophobic pocket are attained (residues Thr7.42, Leu6.51, Leu3.33 and Thr3.36). Nevertheless, these additional hydrophobic interactions do not contribute to an enhancement of the estimation of the free energy of binding of CPA, while in the case of CCPA the combination of the N6 and 2-Cl substitutions leads to a slight increase of predicted affinity with respect to the natural agonist ( $\Delta\Delta G = 0.8$  kcal mol<sup>-1</sup>).

As was also the case with the AUTODOCK scoring function, the LIE predictions of the free energy of binding are much better for the other considered binding position (position B). The potency profile of the four ligands studied is now close to the observed order of potency (see Table 3), although the infra predicted value for the binding of CPA results in an inverted order of the pair NECA-CPA. Analysis of the MD simulations in more detail suggests an explanation for these results and allows an interpretation of how the ranking of compounds is related to the chemical modifications on the adenosine scaffold. First of all, there is a better arrangement of the ribose in its polar binding site than in the binding position A, adopting in this case the north-*anti* conformation in the three ribose unmodified ligands (Fig. 3A, 3C and 3D). Stronger electrostatic interactions are achieved: both the hydroxyl groups at the 2' and 3' ribose positions show stable H-bonds with Asp2.50, whereas the 3' one interacts as well with Ser7.46 and OH5' displays an alternating H-bond with Ser1.40 and His7.43. In the case of NECA (Fig 3B), the chemical modification on 5' induces changes of this interaction profile: the 5'-amide makes additional H-bonds with the sidechains of Asp2.50 and Ser1.46, while the sidechain of His7.43 changes slightly its conformation due to a stable H-bond with O4'. The ethyl chain of the N-ethylcarboxamido group is located in the groove between helices 1 and 2, surrounded by hydrophobic residues (Ala1.43, Val1.47, Gly2.54 and Ile2.58). This pocket is

well conserved between A<sub>1</sub> and A<sub>2</sub> receptors (the only change is Val1.47Ile in A<sub>2A</sub>), which is in agreement with the non-selective profile of this ligand. Additionally, this pocket shows steric hindrance for large substituents, a fact that could explain the low tolerance for such modifications in the 5'-carboxamido derivatives.<sup>[47]</sup> The different arrangement of this ligand in "position B" results in a stronger binding for NECA than for the related 5' unsubstituted agonist, the natural ligand adenosine, although in this case the ligand NECA adopts the south conformation.

In all cases the adenine core is stabilized by hydrophobic interactions with the hydrophobic sidechains of Val3.32 and Trp6.48. Mutagenesis data available for A<sub>1</sub> and A<sub>2A</sub> ARs reveal that a hydrophobic amino acid is necessary at position 3.32.<sup>[18,48]</sup> Regarding Trp6.48, which is highly conserved in GPCRs, experiments on A<sub>3</sub>AR have shown its critical role in receptor activation,<sup>[49]</sup> which is consistent with results obtained on other GPCRs.<sup>[50,51]</sup> Additional interactions of the adenine core involve one H-bond between the nitrogen in position 1 of the adenine ring and the hydroxyl group of Thr3.36, bridged through a water molecule, whereas the amino group in N6 is located close to Thr7.42. These interactions are maintained in the four complexes through the MD simulations, although slight differences are observed among ligands in the strength of a given contact. Concerning the N6 substitution on the adenosine scaffold, the cyclopentyl group of CPA and CCPA appears surrounded by a hydrophobic pocket formed by Leu6.51, Ile7.39, Thr7.42, and in a minor way Leu3.33 and Thr7.35 (Fig 3C and 3D), a pocket that contributes to an increase of the estimated affinity of both ligands with respect to adenosine in agreement with experimental data (see Table 3). Several pieces of experimental data point towards these residues in the ligand binding process. Mutation of Leu 3.33 to alanine results in a substantial reduction of agonist binding, but also of N-0840, which is an adenine-like antagonist bearing a cyclopentyl group in the same position as CPA.<sup>[18]</sup> It is also worth noting that CPA exhibits more than 10,000-fold higher affinity for hA<sub>1</sub> than for hA<sub>2B</sub>,<sup>[12,52]</sup> a related receptor where Thr7.42 is replaced by a serine and Leu3.33 is replaced by a valine. We presume that this affinity difference could be related with an optimal complementarity, and consequently improved hydrophobic interactions, between the N6-substituent and the corresponding receptor pocket of the A<sub>1</sub> receptor. Regarding the nature of the residue 7.35, it has been implicated in species differences in the binding of N6-substituted adenine analogs.<sup>[53]</sup>

The calculated LIE binding free energies predict the CPA complex in binding position B to be more than 2 kcal mol<sup>-1</sup> more stable than the corresponding one with ADO (see Table 2). The interactions between the shared adenosine scaffold (ribose and adenine) and the receptor are the same in both molecules and consequently the N6-cyclopentyl group appears responsible for the enhancement of binding affinity for CPA in comparison to ADO.

If we look at the influence of the 2-chloro substitution in binding position B, the ligand bearing such a chemical modification (CCPA) is predicted to bind stronger than the corresponding ligand without the chlorine, CPA ( $\Delta\Delta G = 2.1$  kcal mol<sup>-1</sup>). This difference is in qualitative agreement with the experimental difference of 0.6 kcal mol<sup>-1</sup> reported by Klotz et al,<sup>[12]</sup> but clearly overestimated here (chlorine charges resulting from the RESP procedure may, however, apparently be more uncertain than those for other atoms<sup>[54]</sup>). The chlorine atom of CCPA is favoring the anchoring of the ligand into the binding site, making slightly favorable electrostatic interactions with Thr3.36, while the interactions of the common scaffold of both ligands are conserved. On the other hand, the main difference between these two ligands comes from the behavior in water of CCPA, that alternates frequently between the south conformation and the north conformation, while CPA does not explore the south conformation in water. This conformational equilibrium results in a ligand electrostatic interaction energy ( $\langle V_{l-s}^{el} \rangle_w$ ) that is less negative for CCPA, with a concomitant increase of its predicted affinity.

NECA would bind to the hA<sub>1</sub>AR stronger than ADO ( $\Delta\Delta G_{\text{bind}} = -3.1$  kcal mol<sup>-1</sup>) and 1.9 kcal mol<sup>-1</sup> weaker than CCPA, again in qualitative agreement with the experimental differences. On the other hand, our results predict essentially the same binding energy for NECA than for CPA. Thus, only for this pair of ligands the experimental ranking is not reproduced, since there is an experimental difference of 1.1 kcal mol<sup>-1</sup> in favor of CPA. It should be noted here that the predicted binding energy of NECA is associated with larger convergence errors than the other compounds, which makes the comparison for this particular pair of molecules more uncertain.

## CONCLUDING REMARKS

The previously reported hA<sub>1</sub>AR putative agonist binding site has been investigated by means of docking and binding free energy predictions of representative agonists, in an attempt to reproduce in our molecular model the known

structure-activity relationships of A<sub>1</sub>AR agonists.<sup>[6]</sup> In the first part of this study, an exhaustive docking exploration of receptor model with the four agonists considered was performed, using the AUTODOCK software. It was aimed at covering all available possibilities for agonist-receptor docking and yielded the result that no other solution apart from those denoted "A" and "B" should be considered for further investigation. Regarding these two docking positions, the empirical scoring function implemented in the docking software pointed to position B as the most predictive of experimental affinity data. Additionally, the superposition of the four ligands on their predicted position B was much better than on the alternative position A, which was only found for three of them.

The analysis of both docking alternatives with the MD/LIE approach dealt with a double objective: first, to provide a detailed description of the ligand-receptor interactions in a more realistic, dynamic environment, with the aim of better identifying the residues involved in each considered binding mode and the effects of the considered chemical modifications of adenosine. MD simulations allow the relaxation of the receptor-ligand complexes, and therefore the resulting pattern of specific interactions may be different from the initial static picture of a docked complex. In fact, some complexes needed an equilibration period after which the average energies were stabilized, reflecting that small conformational changes can lead to increased ligand-receptor affinities. The second objective was to assess the docking results through estimation of the ligand binding free energy obtained with the LIE method. This kind of assessment of docking results with the LIE approach was found useful in the investigation of the binding of blockers to K<sup>+</sup> channels by Luzhkov and Åqvist.<sup>[55]</sup> In our case, the LIE method predicts more stable complexes for the binding position B in all cases, with binding energy differences between position A and B that go from -0.8 to -9.2 kcal mol<sup>-1</sup> (extreme values for ADO and NECA, respectively). The other issue that arises is whether the affinity variations between compounds are correctly predicted in position B and what are the structural reasons for these differences.

A comparison between the estimated (with both the AUTODOCK scoring function and LIE approaches) and experimental free energies of binding for position B is depicted in Fig. 4. Here, the absolute values derived from the LIE approach have been adjusted to experimental values with a least-squares derived constant parameter  $\gamma = -7.3$  kcal mol<sup>-1</sup>, that is found to be identical to that earlier obtained for retinol binding protein.<sup>[33]</sup> Although the scope of the present work is not the prediction of absolute binding free

energies, the comparison of the differential values obtained is facilitated by the inclusion of this parameter in the LIE equation. As discussed above, the interpretation of the  $\gamma$  parameter is related to binding site hydrophobicity and possibly to neglected features in our homology derived model, i.e. absence of loop regions and membrane environment, that should affect all agonists in the same way. The origin of a negative  $\gamma$  for hydrophobic binding sites has been addressed elsewhere<sup>[45]</sup> and can be understood from the extrapolations of hydrophobic solvation energies to zero solute size in polar and nonpolar environments. For a discussion about the relevant thermodynamic cycles describing the microscopic binding process, see also ref. [56]. Fig 4 shows that the ranking within the series is quite reasonable with both AUTODOCK and LIE approaches, even if the ligand CPA is overestimated by the scoring function and underestimated by the LIE approach. Moreover, the role of the chemical substitutions that enhance the affinity with respect to the unmodified adenosine can be explained in terms of ligand-receptor interactions. A hydrophobic pocket is identified between helices 3, 6 and 7 for bulky N6 substitutions characteristic of A<sub>1</sub> selective agonists; the 2-chloro substitution, present in the most potent agonist of the series, is well accommodated between helices 3 and 6; finally, the 5' substitution of 5'-N-alkylcarboxamidoadenosines is located in a binding cavity with limited available space between helices 1, 2 and 7, thus not suitable for large N-alkyl modifications. Additionally, the proposed model for agonists binding to the hA<sub>1</sub>AR is explanatory for some mutagenesis data that affect all the compounds in the series, such as those related to Val3.32,<sup>[48]</sup> Thr3.36<sup>[18]</sup> and His7.43.<sup>[19,20]</sup> Our binding mode is also consistent with the possible implications of Trp6.48 on receptor activation,<sup>[49]</sup> although this biological process can not be simulated with the current methodology.

Recently, a model of agonist-receptor interactions has been published for the A<sub>2A</sub> subtype,<sup>[57]</sup> in which the binding site proposed partially overlaps the one postulated in the present study, although differences are also observed. In addition to the fact that both studies have been carried out in related but different receptors (hA<sub>2A</sub>AR vs. hA<sub>1</sub>AR), it should be pointed out that different strategies underlie the two studies. The main aim of the study of Kim and co-workers<sup>[57]</sup> was to generate a model in agreement with the experimental evidence. With this purpose the authors introduced manual interventions into the modeling process. Conversely, the present study uses automatic protocols free of human intervention to obtain ligand-protein models.

Experimental evidence are then used a posteriori to choose among different solutions and to assess the quality of the models. Taking into account the complexity of the interpretation of mutagenesis data<sup>[58]</sup> the two modeling approaches are, in our opinion, complementary and useful for dealing with the prediction of ligand binding positions in these systems. Besides the relevance of the present modeling results for guiding the design of more potent agonists of the A<sub>1</sub>AR, it should be pointed out that the described strategy consisting of obtaining docking positions automatically and assessing them by checking their consistency for a series of ligands, could be useful in other similar challenging situations.

#### **ACKNOWLEDGEMENTS**

The authors thank Almirall Prodesfarma S.A. and Dr. Mabel Loza for their information and support. They are also grateful to Dr. Jordi Villà-Freixa for helpful discussions and careful reading of the manuscript. HGT acknowledges CIRIT of Generalitat de Catalunya for his PhD grant.

#### **REFERENCES**

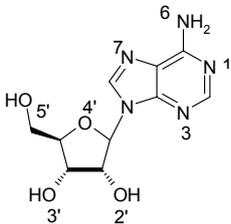
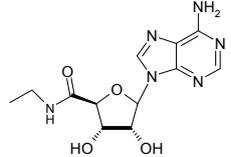
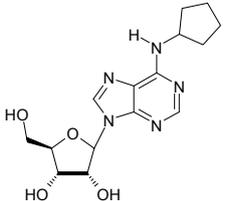
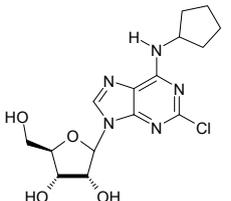
1. C.E. Müller, *Farmaco* **2001**, *56*, 77-80.
2. C. Müller, B. Stein, *Curr. Pharm. Design.* **1996**, *2*, 501-530.
3. S.A. Poulsen, R.J. Quinn, *Bioorg. Med. Chem.* **1998**, *6*, 619-641.
4. H.L. Haas, O. Selbach, *Naunyn Schmiedebergs Arch. Pharmacol.* **2000**, *362*, 375-381.
5. J.M. Downey, G.S. Liu, J.D. Thornton, *Cardiovasc. Res.* **1993**, *27*, 3-8.
6. C.E. Müller, *Curr. Med. Chem.* **2000**, *7*, 1269-1288.
7. S. Vittori, A. Lorenzen, C. Stannek, S. Costanzi, R. Volpini, I.J. AP, J.K. Kunzel, G. Cristalli, *J. Med. Chem.* **2000**, *43*, 250-260.
8. E.M. van der Wenden, J.K. von Frijtag Drabbe Kunzel, R.A. Mathot, M. Danhof, I.J. AP, W. Soudijn, *J. Med. Chem.* **1995**, *38*, 4000-4006.
9. M.J. Lohse, K.N. Klotz, E. Diekmann, K. Friedrich, U. Schwabe, *Eur. J. Pharmacol.* **1988**, *156*, 157-160.
10. J.W. Daly, W.L. Padgett, *Biochem. Pharmacol.* **1992**, *43*, 1089-1093.

11. M.J. Lohse, K.N. Klotz, U. Schwabe, G. Cristalli, S. Vittori, M. Grifantini, *Naunyn Schmiedebergs Arch. Pharmacol.* **1988**, 337, 687-689.
12. K.N. Klotz, J. Hessling, J. Hegler, C. Owman, B. Kull, B.B. Fredholm, M.J. Lohse, *Naunyn Schmiedebergs Arch. Pharmacol.* **1998**, 357, 1-9.
13. K.N. Klotz, *Naunyn Schmiedebergs Arch. Pharmacol.* **2000**, 362, 382-391.
14. H. Gutiérrez-de-Terán, N.B. Centeno, M. Pastor, F. Sanz, *Proteins*, in press.
15. K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, M. Miyano, *Science* **2000**, 289, 739-745.
16. P.J. Goodford, *J. Med. Chem.* **1985**, 28, 849-857.
17. K. Kristiansen, W.K. Kroeze, D.L. Willins, E.I. Gelber, J.E. Savage, R.A. Glennon, B.L. Roth, *J. Pharmacol. Exp. Ther.* **2000**, 293, 735-746.
18. S.A. Rivkees, H. Barbhaiya, A.P. IJzerman, *J. Biol. Chem.* **1999**, 274, 3617-3621.
19. K. Klotz, M. Lohse, U. Schwabe, *J. Biol. Chem.* **1988**, 263, 17522-17526.
20. M.E. Olah, H. Ren, J. Ostrowski, K.A. Jacobson, G.L. Stiles, *J. Biol. Chem.* **1992**, 267, 10764-10770.
21. J.A. Ballesteros, H. Weinstein in *Methods in Neurosciences* (Eds.: P.M. Conn, S.C. Sealfon), Academic Press, San Diego, **1994**, pp. 366-428.
22. F.H. Allen, O. Kennard, *Chem. Des. Autom. News.* **1993**, 8, 31-37.
23. Insight II, Molecular simulations Inc., San Diego, **2000**.
24. J.J. Stewart, *J. Comput. Aided. Mol. Des.* **1990**, 4, 1-105.
25. Gaussian 98, Gaussian Inc., Pittsburgh, PA, USA, **1998**.
26. C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman, *J. Phys. Chem.* **1993**, 97, 10269-10280.
27. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, *J. Am. Chem. Soc.* **1995**, 117, 5179-5197.
28. G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, *J. Comp. Chem.* **1998**, 19, 1639-1662.

29. J. Åqvist, C. Medina, J.E. Samuelsson, *Protein. Eng.* **1994**, 7, 385-391.
30. T. Hansson, J. Marelius, J. Åqvist, *J. Comput. Aided. Mol. Des.* **1998**, 12, 27-35.
31. F.S. Lee, Z.T. Chu, M.B. Bolger, A. Warshel, *Prot. Eng.* **1992**, 5, 215-228.
32. J. Åqvist, T. Hansson, *J. Phys. Chem.* **1996**, 100, 9512-9521.
33. J. Åqvist, J. Marelius, *Comb. Chem. High Throughput. Screen.* **2001**, 4, 613-626.
34. J. Marelius, K. Kolmodin, I. Feierberg, J. Åqvist, *J. Mol. Graph. Modelling* **1999**, 16, 213-225.
35. W. Jorgensen, J. Chandrasekhar, J. Madura, I. RW, M. Klein, *J. Chem. Phys.* **1983**, 79, 926-935.
36. J. Essex, W. Jorgensen, *J. Comp. Chem.* **1995**, 16, 951-972.
37. G. King, A. Warshel, *J. Chem. Phys.* **1989**, 91, 3647-3661.
38. G. Allende, V. Casado, J. Mallol, R. Franco, C. Lluís, E.I. Canela, *J. Neurochem.* **1993**, 60, 1525-1533.
39. F.S. Lee, A. Warshel, *J. Chem. Phys.* **1992**, 97, 3100-3107.
40. W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph. Model.* **1996**, 14, 33-38.
41. E. Van der Wenden, S.L. Price, R.P. Apaya, A.P. IJzerman, W. Soudijn, *J. Comp. Aid. Mol. Des.* **1995**, 9, 44-54.
42. M. Rashid, P. Manivet, H. Nishio, J. Pratuangdejkul, M. Rajab, M. Ishiguro, J.M. Launay, T. Nagatomo, *Life Sci.* **2003**, 73, 193-207.
43. M. Mahmoudian, *J. Mol. Graph. Model.* **1997**, 15, 149-153, 179.
44. A. Gieldon, R. Kazmierkiewicz, R. Slusarz, J. Ciarkowski, *J. Comput. Aided. Mol. Des.* **2001**, 15, 1085-1104.
45. M. Almlöf, B.O. Brandsdal, J. Åqvist, *submitted for publication*.
46. P. Franchetti, L. Cappellacci, S. Marchetti, L. Trincavelli, C. Martini, M.R. Mazzoni, A. Lucacchini, M. Grifantini, *J. Med. Chem.* **1998**, 41, 1708-1715.
47. M. de Zwart, A. Kourounakis, H. Kooijman, A.L. Spek, R. Link, J.K. von Frijtag Drabbe Kunzel, I.J. AP, *J. Med. Chem.* **1999**, 42, 1384-1392.

48. Q. Jiang, B.X. Lee, M. Glashofer, A.M. van Rhee, K.A. Jacobson, *J. Med. Chem.* **1997**, *40*, 2588-2595.
49. Z.-G. Gao, A. Chen, D. Barak, S.-K. Kim, C.E. Muller, K.A. Jacobson, *J. Biol. Chem.* **2002**, *277*, 19056-19063.
50. J.A. Javitch, J.A. Ballesteros, H. Weinstein, J. Chen, *Biochemistry* **1998**, *37*, 998-1006.
51. J. Marie, E. Richard, D. Pruneau, J.L. Paquet, C. Siatka, R. Larguier, C. Ponce, P. Vassault, T. Groblewski, B. Maigret, J.C. Bonnafous, *J Biol Chem* **2001**, *276*, 41100-41111.
52. J. Linden, T. Thai, H. Figler, X. Jin, A.S. Robeva, *Mol. Pharmacol.* **1999**, *56*, 705-713.
53. A.L. Tucker, A.S. Robeva, H.E. Taylor, D. Holeton, M. Bockner, K.R. Lynch, J. Linden, *J. Biol. Chem.* **1994**, *269*, 27900-27906.
54. M. Almlöf, *Personal communication*.
55. V.B. Luzhkov, J. Aqvist, *FEBS Lett.* **2001**, *495*, 191-196.
56. Y. Sham, Z.T. Chu, H. Tao, A. Warshel, *Proteins* **2000**, *39*, 393-407.
57. S.-K. Kim, Z.-G. Gao, P.V. Rompaey, A.S. Gross, A. Chen, S.V. Calenbergh, K.A. Jacobson, *J. Med. Chem.* **2003**, *46*, 4847-4859
58. D. Colquhoun, *Br. J. Pharmacol.* **1998**, *125*, 924-947.
- .

**Table 1** $A_1$  adenosine receptor agonists studied.

Compound <sup>[a]</sup>	Name	$A_1$ AR $K_i$ (nM)	$\Delta G^{[d]}$ (kcal mol <sup>-1</sup> )
	ADO	73 <sup>[b]</sup>	-9.8
	NECA	13.6 <sup>[c]</sup>	-10.8
	CPA	2.25 <sup>[c]</sup>	-11.9
	CCPA	0.83 <sup>[c]</sup>	-12.5

[a] Standard numbering of adenosine derivatives is shown for adenosine.

[b] Inhibition of adenylate cyclase in rat  $A_1$ AR.<sup>[10]</sup>

[c] Values in human  $A_1$ AR extracted from ref. [12].

[d]  $\Delta G$  calculated using  $K_i = e^{\Delta G/RT}$  at room temperature (25°C).

**Table 2**

Summary of the docking results for the four A<sub>1</sub>AR agonists, showing the energy of binding predicted by the scoring function and the ranking of each solution among those provided by AUTODOCK.

Compound	Position A		Position B	
	$\Delta G$	Rank	$\Delta G$	Rank
ADO	-10.6	1 <sup>[a]</sup>	-10.4	3 <sup>[a]</sup>
NECA	-	-	-11.1	1
CPA	-10.7	6	-13.5	1
CCPA	-10.8	6	-12.7	1

<sup>[a]</sup> Docking solutions 1-3 for ADO are, in practice, energetically equivalent.

**Table 3**

Average ligand-surrounding interaction energies (kcal/mol) for each compound in water and in the two binding positions are given together with the calculated relative binding free energies.

Ligand	Water		Position A			Position B		
	$\langle V_{l-s}^{el} \rangle_w$	$\langle V_{l-s}^{vdW} \rangle_w$	$\langle V_{l-s}^{el} \rangle_p$	$\langle V_{l-s}^{vdW} \rangle_p$	$\Delta\Delta G_{bind}^{calc}$ [a]	$\langle V_{l-s}^{el} \rangle_p$	$\langle V_{l-s}^{vdW} \rangle_p$	$\Delta\Delta G_{bind}^{calc}$ [a]
ADO	-69.4	-16.5	-61.7	-33.0	-0.5±0.4	-63.3	-34.8	-1.3±0.6
NECA	-76.1	-20.2	-49.6	-41.9	4.8±2.1	-77.1	-42.7	-4.4±1.4
CPA	-65.0	-25.7	-46.1	-50.7	2.1±0.5	-65.1	-48.4	-4.2±0.4
CCPA	-58.3	-26.5	-48.8	-50.7	-1.3±0.9	-65.1	-47.5	-6.3±0.5

[a] Binding free energies calculated from Eq. (1) without any optimization of the constant  $\gamma$ .

## FIGURE CAPTIONS

**Fig. 1:** Superposition of agonists considered in the docking positions A (left) and B (right) obtained by AUTODOCK. Ligands are colored in red (ADO), blue (CPA), green (CCPA) and yellow (NECA). Note that this last ligand was not found in position A (see text for details).

**Fig. 2:** South and North conformations of adenosine. The dashed line indicates the intramolecular hydrogen bond between OH2' and N3, present in the south conformation.

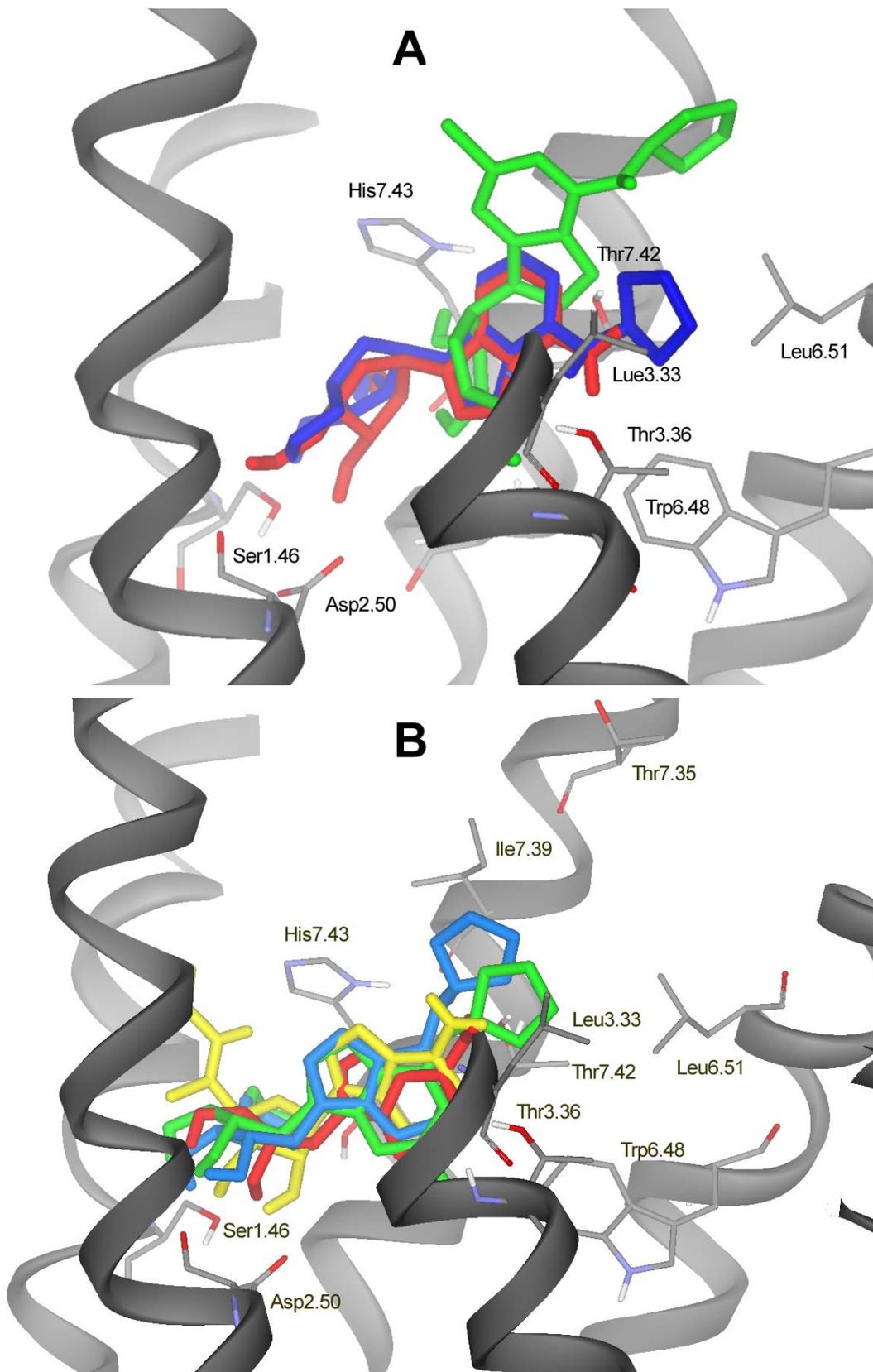
**Fig. 3:** Snapshots from MD simulations of ADO (**A**), NECA (**B**) CPA (**C**) and CCPA (**D**) in binding position B. Dashed lines indicate hydrogen bonds frequently detected in the MD trajectory.

**Fig. 4:** Comparison of experimental (OBS), AUTODOCK and LIE estimated free energies of binding for the binding position B. Standard errors of experimental energies are derived from values in the original references.

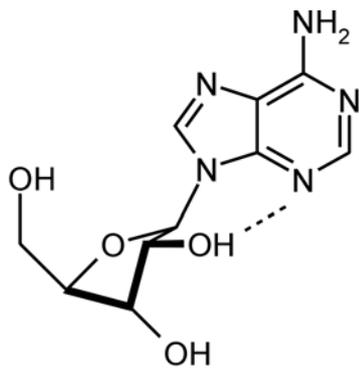
**Text for the table of contents:**

Superposition of four adenosine derivatives docked to an A1 adenosine receptor model. Docking, molecular dynamics and free energy calculations have been used in this study to estimate the relevance of the proposed ligand-receptor interactions.

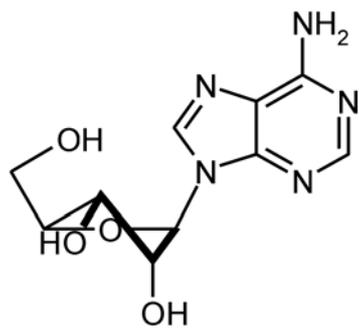
Figure 1



**Figure 2**



South



North

Figure 3

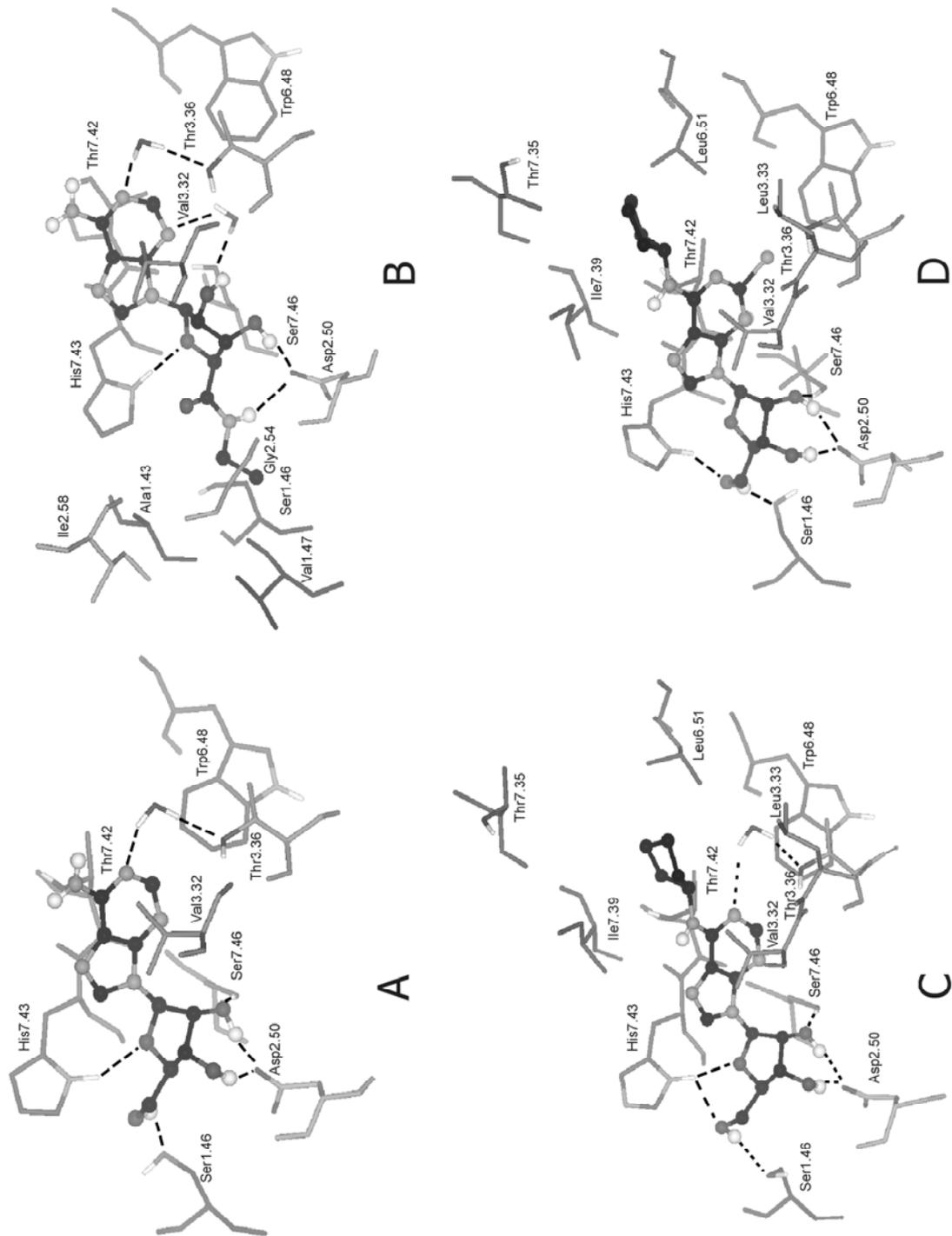
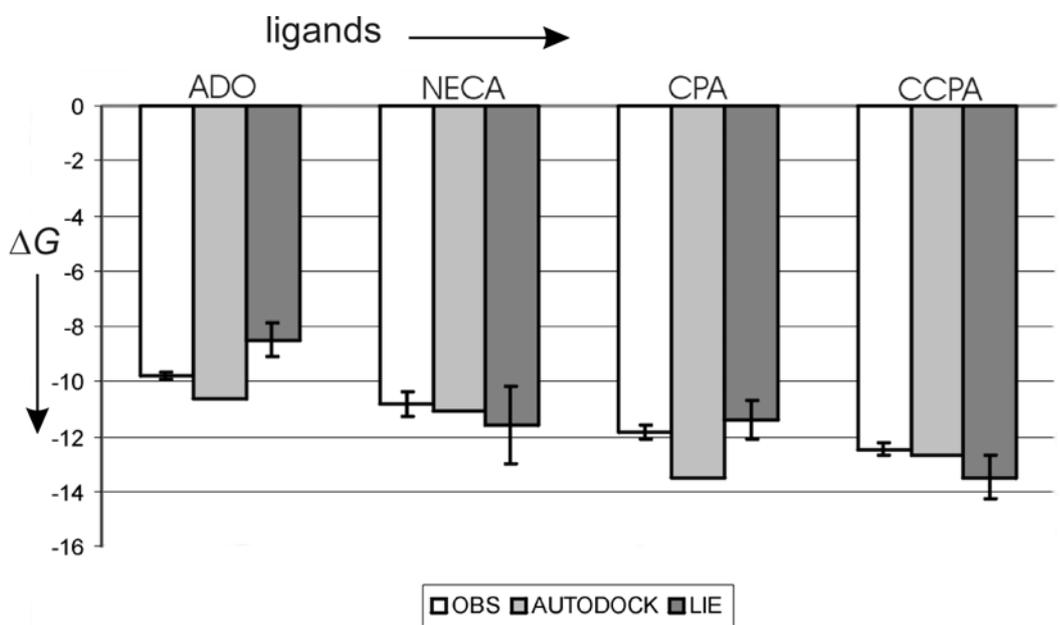
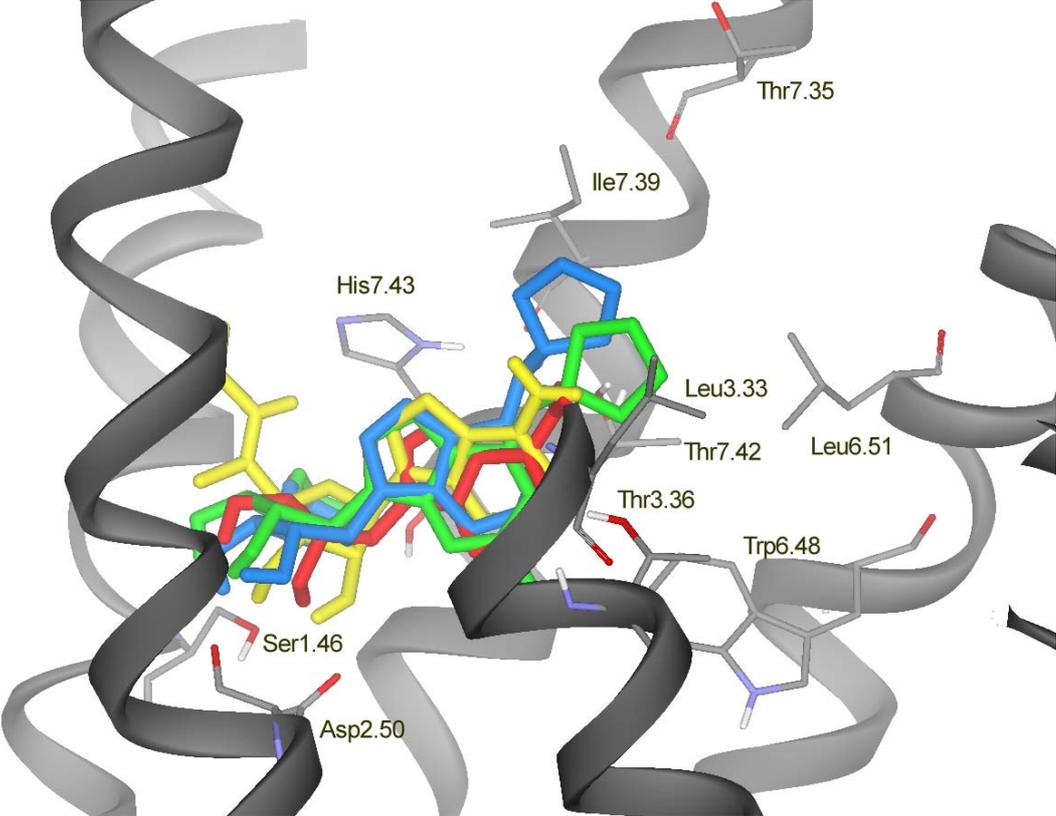


Fig 4



**Table of contents figure**



## APÉNDICES



- A1** **Script** para la para la automatización del cálculo de propiedades químicas utilizado en el Artículo I.
- A2** **Resumen del Artículo** "Use of alignment-free molecular descriptors in diversity analysis and optimal sampling of molecular libraries."
- A3** **Poster** presentado en el **XII Congreso de la Sociedad Española de Química Terapéutica.**
- A4** **Resumen del Artículo** "On the Generation of Catalytic Antibodies by Transition State Analogues."
- A5** **Resumen del Artículo** "Towards a MIP-based alignment and docking in computer-aided drug design."

```

#!/bin/csh
ls *.mol | sed 's/.mol//' > list.lst
foreach name (`cat list.lst`)
corina -ttracefile=$name.trc -it=sdf $name.mol -ot=mol2 $name.mol2
#generates 3D structures
babel -imol2 "$name".mol2 -omopint "$name".dat "AM1 SM5.4A SOLVNT=WATER"
#generates internal coord.
AMSOL "$name"
#launches AMSOL calculation
# Getting calculated values
grep " Dipole moment from CM1 point charges" = "$name".arc | gawk '\
    BEGIN{nom=ARGV[1];ARGV[1]="";}{print nom,$8}\
    ' $name >> DIPOLE-MOMENT
grep "(most abundant/longest-lived isotopes)" = "$name".arc | gawk '\
    BEGIN{nom=ARGV[1];ARGV[1]="";}{print nom,$5}\
    ' $name >> MOLECULAR-WEIGHT
grep "Total:" "$name".arc | gawk '\
    BEGIN{nom=ARGV[1];ARGV[1]="";}{print nom,$4}\
    ' $name >> SURFACE-AREA
grep "Total:" "$name".arc | gawk '\
    BEGIN{nom=ARGV[1];ARGV[1]="";}{print nom,$6}\
    ' $name >> AG=WATER

gawk '\
{if ($3=="0" && NF==12) bit=1;if ($1=="0") bit=0;if (bit) print $0}\
' "$name".arc > "$NAME".tmp
cut -c1-60 "$NAME".tmp > "$NAME"_opt.dat

# benzene calculations

sed 's/SOLVNT=WATER/1SCF SOLVNT=BENZENE/' "$name"_opt.dat > "$name"_bz.dat
AMSOL "$name"_bz
grep "Total:" "$name"_bz.arc | gawk '\
    BEGIN{nom=ARGV[1];ARGV[1]="";\
    }{print nom,$6}' "$name"_bz >> AG=Bz

# Octanol calculations

sed 's/SOLVNT=BENZENE/SOLVNT=GENORG' "$name"_bz.dat > tmp.dat
cat <EOF>> "$name"_oct.dat
IOFR=1.4295 ALPHA=0.37 BETA=0.48 GAMMA=39.0
&DIELEC=9.87 FACARB=0.00 FEHALO=0.0
EOF

cat tmp.dat >> "$name"_oct.dat
AMSOL "$name"_oct
grep "Total:" "$name"_oct.arc | gawk '\
    BEGIN{nom=ARGV[1];ARGV[1]="";\
    }{print nom,$6}' "$name"_oct >> AG=oct

# Chloroform calculations

sed 's/SOLVNT=BENZENE/SOLVNT=CHCL3' "$name"_bz.dat > "$name"_chcl3.dat
grep "Total:" "$name"_chcl3.arc | gawk '\
    BEGIN{nom=ARGV[1];ARGV[1]="";\
    }{print nom,$6}' "$name"_chcl3 >> AG=chcl3
end

paste DIPOLE-MOMENT SURFACE-AREA > tmp1
paste MOLECULAR-WEIGHT AG=water > tmp2
paste AG=oct AG=chcl3 > tmp4
paste tmp1 tmp2 > tmp3
paste tmp3 tmp4 > variables.dat

rm *tmp*

echo "Your data have been stored into file variables.dat"

```



*Full Paper*

## Use of alignment-free molecular descriptors in diversity analysis and optimal sampling of molecular libraries

Fabien Fontaine, Manuel Pastor, Hugo Gutiérrez-de-Terán, Juan J. Lozano & Ferran Sanz\*

*Research Group on Biomedical Informatics (GRIB), IMIM, Universitat Pompeu Fabra, C/ Dr. Aiguader 80, Barcelona, Spain*

(\* Author for correspondence, E-mail: fsanz@imim.es, Fax: +34 932 240 875)

Received 5 May 2003; Accepted 11 July 2003

**Key words:** Almond, diversity, GRIND, k-means clustering, molecular descriptors, molecular library sampling, principal component analysis, quantum-mechanical descriptors, VolSurf

### Summary

The selection of a sample of diverse compounds is a common strategy for exploring large molecular libraries. However, the success of such approach depends on the selection of relevant molecular descriptors and the use of appropriate sampling methods. In the context of pharmaceutical research, the molecular descriptors should be based on physicochemical properties related with the pharmacological behaviour of the compounds. In this sense, the alignment-free GRIND and VolSurf molecular descriptors are promising candidates since they have been successfully used in the modelling of both pharmacodynamic and pharmacokinetic properties of drugs. This work describes the use of such descriptors in the diversity sampling of a library of primary amines and compares the results with those obtained in a previous study that used quantum-mechanical descriptors. As in the previous work, principal component (PC) analysis was applied to reduce the dimensionality and remove redundant information of the original descriptors, and the compounds were sampled on the basis of k-means clustering on the space of the selected PCs. The results of the present study show that VolSurf and GRIND provide similar quality sampling regarding global features of the molecules such as hydrophilicity, however the topology of the compounds is considered differently. The similarity between particular compounds strongly depends on the original descriptors used. However all the sample selections done in the PC space after k-means clustering provide the same apparent diversity in comparison to the whole dataset. The results indicate that there is no best set of descriptors on a diversity basis. The selection of descriptors must be based on the drug features to be investigated.

**Abbreviations:** 2-D: 2-Dimensional; 3-D: 3-Dimensional; GRIND: GRid-INdependent Descriptors; MACC2: Maximum Auto- and Cross-Correlation; MIF: Molecular Interaction Field; PC: Principal Component; PCA: Principal Component Analysis; QM: Quantum-Mechanical; QSAR: Quantitative Structure-Activity Relationships.

# On the Generation of Catalytic Antibodies by Transition State Analogues

Montserrat Barbany,<sup>[a]</sup> Hugo Gutiérrez-de-Terán,<sup>[a]</sup> Ferran Sanz,<sup>\*,[a]</sup>  
Jordi Villà-Freixa,<sup>\*,[a]</sup> and Arieh Warshel<sup>\*,[b]</sup>

*The effective design of catalytic antibodies represents a major conceptual and practical challenge. It is implicitly assumed that a proper transition state analogue (TSA) can elicit a catalytic antibody (CA) that will catalyze the given reaction in a similar way to an enzyme that would evolve (or was evolved) to catalyze this reaction. However, in most cases it was found that the TSA used produced CAs with relatively low rate enhancement as compared to the corresponding enzymes, when these exist. The present work explores the origin of this problem, by developing two approaches that examine the similarity of the TSA and the corresponding transition state (TS). These analyses are used to assess the proficiency of the CA generated by the given TSA. Both approaches focus on electrostatic effects that have been found to play a major role in enzymatic reactions. The first method uses molecular interaction potentials to look for the similarity between the TSA and the TS and, in principle, to help in designing new haptens by using 3D quantitative structure–activity relationships. The second and more quantitative approach generates a grid of Langevin*

*dipoles, which are polarized by the TSA, and then uses the grid to bind the TS. Comparison of the resulting binding energy with the binding energy of the TS to the grid that was polarized by the TS provides an estimate of the proficiency of the given CA. Our methods are used in examining the origin of the difference between the catalytic power of the 1F7 CA and chorismate mutase. It is demonstrated that the relatively small changes in charge and structure between the TS and TSA are sufficient to account for the difference in proficiency between the CA and the enzyme. Apparently the environment that was preorganized to stabilize the TSA charge distribution does not provide a sufficient stabilization to the TS. The general implications of our findings and the difficulties in designing a perfect TSA are discussed. Finally, the possible use of our approach in screening for an optimal TSA is pointed out.*

## KEYWORDS:

catalytic antibodies · Langevin dipoles · molecular interaction potentials · structure–activity relationships · transition states

[a] Prof. F. Sanz, Dr. J. Villà-Freixa, M. Barbany, H. Gutiérrez-de-Terán  
Computational Structural Biology Laboratory  
Research Group on Biomedical Informatics (GRIB)—IMIM/UPF  
Passeig Marítim de la Barceloneta 37–49  
08003 Barcelona (Spain)  
Fax: (+34) 93-224-0875  
E-mail: fsanz@imim.es, jvilla@imim.es

[b] Prof. A. Warshel  
Department of Chemistry  
University of Southern California  
Los Angeles, CA 90089-1062 (USA)  
Fax: (+1) 213-740-2701  
E-mail: warshel@usc.edu

# SUPERPOSICIONES ESTRUCTURALES NO EVIDENTES A PARTIR DE DISTRIBUCIONES DE PEM. APLICACIÓN EN ANTAGONISTAS DEL RECEPTOR DE ADENOSINA A<sub>2A</sub>

Hugo G. de Terán,<sup>a</sup> Nuria B. Centeno,<sup>a</sup> Victor Segarra,<sup>b</sup> y Ferran Sanz<sup>a</sup>

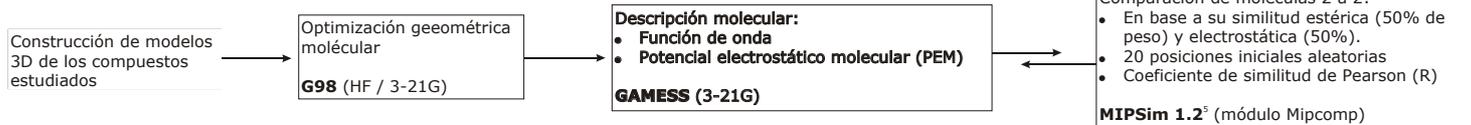
<sup>a</sup>G.R. En Informàtica Biomèdica, IMIM/UPF, Dr. Aiguader, 80, 08003 Barcelona.

<sup>b</sup>Almirall Prodesfarma, Centro de Investigación, Cardener, 68-74, 08024 Barcelona.

## INTRODUCCIÓN

La señalización intracelular generada por la adenosina extracelular está mediada por una serie de receptores de membrana (A<sub>1</sub>, A<sub>2A</sub>, A<sub>2B</sub> y A<sub>3</sub>).<sup>1</sup> Consecuentemente, existe un interés creciente en desarrollar herramientas farmacológicas (ligandos selectivos y de alta afinidad) para estudiar estos receptores y establecer sus funciones fisiológicas. Un aspecto crucial para el diseño de estas herramientas es la definición de sus requerimientos estéricos y electrostáticos. Las moléculas que se unen a los receptores de adenosina pertenecen a distintas familias químicas,<sup>2</sup> que no muestran semejanzas estructurales evidentes.

## MÉTODOS

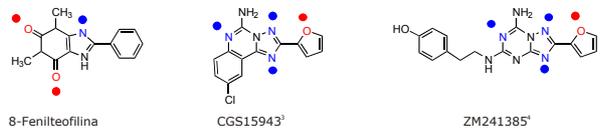


## RESULTADOS

Cada una de las tres familias estudiadas tiene tres sustituyentes sobre su sistema heterocíclico y un mínimo de tres zonas de aceptación de puente de hidrógeno. Así pues, en cada una de las tres comparaciones existen teóricamente 6 posibles superposiciones que hagan coincidir sustituyentes y aceptores de puentes de hidrógeno. De éstas, mostramos aquellas que poseen un coeficiente de similitud (R) mayor que 0.4

## OBJETIVOS

Usando representantes de tres clases estructurales de antagonistas A<sub>2A</sub> (xantinas, triazolopirimidinas y triazololo[1,5-a][1,3,5]triazinas):



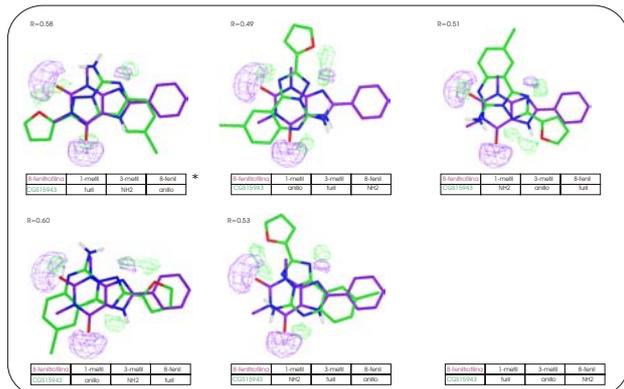
● Puntos de aceptación de puentes de hidrógeno

se pretende:

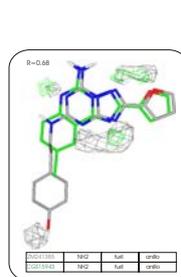
- Explorar las semejanzas entre los tres tipos de antagonistas en base a sus distribuciones de potencial electrostático molecular (PEM).
- Proponer superposiciones estructurales adecuadas.

Las dos soluciones propuestas se basan en el consenso de las tres comparaciones por pares de compuestos, teniendo en cuenta que sólo existe un alineamiento posible entre ZM241385 y CGS15943.

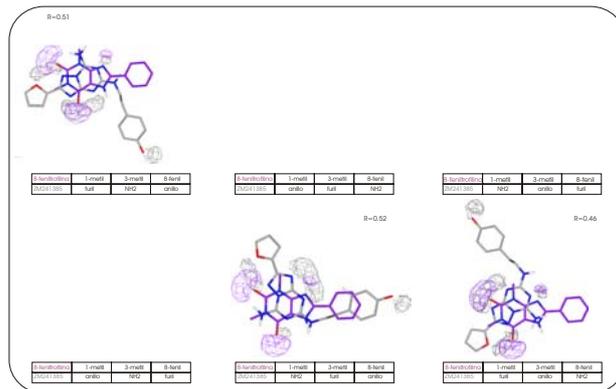
### 8-FENILTEOFILINA vs CGS15943



### ZM241385 vs CGS15943

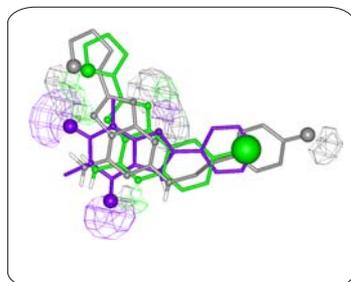


### 8-FENILTEOFILINA vs ZM241385

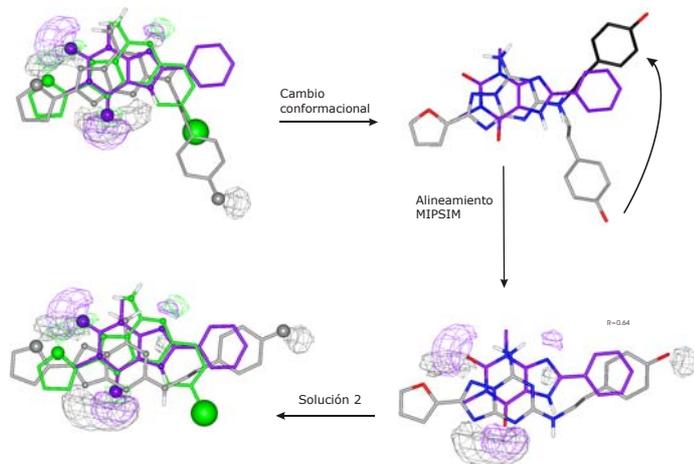


\* correspondencia entre sustituyentes

### SOLUCIÓN 1



En la solución 2 no se observa una buena superposición del *p*-hidroxifeniletilamino del compuesto ZM241385 con respecto a fragmentos cíclicos de los otros dos compuestos. Debido a la libertad conformacional del *p*-hidroxifeniletilamino, hemos realizado un cambio conformacional energéticamente accesible ( $\Delta E=2.5$  Kcal/mol) para superponerlo con el fenilo de la xantina. Una nueva comparación de las dos moléculas con MIPSim nos da una solución con un elevado coeficiente de similitud (R=0.64).



### SOLUCIÓN 2

## BIBLIOGRAFÍA

- (1) Nyce, J.W. *Trends Pharmacol. Sci.* **1999**, *20*, 79.
- (2) Müller, C.E.; Stein, B. *Curr. Pharm. Des.* **1996**, *2*, 501.
- (3) Francis, J.E. *et al.* *J. Med. Chem.* **1988**, *31*, 1014.
- (4) Palmer, T.M.; Poucher S.M.; Jacobson K.A.; Stiles G.L. *Mol. Phar.* **1995**, *48*, 970.
- (5) de Caceres, M.; Villà, J.; Lozano, J.J.; Sanz, F.; *Bioinformatics.* **2000**, *16*, 568.

## AGRADECIMIENTOS

El presente trabajo ha sido financiado parcialmente por la CICYT, el FIS y la CIRIT.

*Last updated November 4, 2003*

## **Towards a MIP-based alignment and docking in computer-aided drug design.**

**Montserrat Barbany<sup>‡</sup>, Hugo Gutiérrez de Terán<sup>‡</sup>, Ferran Sanz\*, and Jordi  
Vilà-Freixa\***

Address:

Research Group on Biomedical Informatics (GRIB) - IMIM/UPF  
Passeig Marítim de la Barceloneta 37-49. Tel: +34 93 224 0886  
E-08003 Barcelona (Spain) Fax: +34 93 224 0875

\*e-mail address: fsanz@imim.es, jvilla@imim.es

<sup>‡</sup>These authors contributed equally to this work.

**ABSTRACT.** Structural alignment of ligands in their biological conformation is a crucial step in the building of pharmacophoric models in structure-based drug design. In addition, docking algorithms are limited in some cases by the quality of the scoring functions and the limited flexibility of the environment that the different programs allow. On the other hand, GRID molecular interaction potentials (MIPs) have been used for a long time in 3D-QSAR studies. However, in most of these studies the alignment of the molecules is performed on the basis of geometrical or physico-chemical criteria that differs from the MIPs used in the partial least squares statistical analysis. We have previously developed a method to use the same scoring function for the molecular alignment and for 3D-QSAR studies. This methodology, based on the use of GRID potentials, consists in the ponderate averaging of similarities of the relevant MIPs of the molecules to be aligned. Here we present a method to obtain the weights of the different GRID probes for the ponderated average based on the structural information on protein-ligand complexes for relevant systems. The method, implemented in MIPSIM, is shown to yield good accuracy in the prediction of the alignments for two systems: a set of three inhibitors of dihydrofolate reductase and a set of fifteen non-nucleoside HIV-1 reverse transcriptase inhibitors (NNRTIs). The smooth GRID potentials are shown to capture the flexible character of the active site, as opposed to traditional docking scoring energy functions.