Universitat Ramon Llull

TESIS DOCTORAL

Título	**Herramientas de cribado virtual aplicadas a inhibidores de entrada del VIH. Diseño de nuevos compuestos anti-VIH.**
Realizada por	Violeta Isabel Pérez Nueno
en el Centro	Escola Tècnica Superior IQS
y en el Departamento	Química Orgánica
Dirigida por	Dr. Jordi Teixidó i Closa, Dr. Xavier Batllori Aguilà

C.Claravall, 1-3
08022 Barcelona
Tel. 936 022 200
Fax. 936 022 249
E-mail: urlsc@sec.url.es
www.url.es

2. Discusión de Resultados

Tal y como se ha expuesto previamente, las terapias anti-HIV actuales presentan dos grandes problemas: la resistencia viral y los efectos secundarios. Por ello, se buscan nuevos fármacos alternativos, con otros modos de acción, que consigan una mayor potencia, menores efectos secundarios y menor riesgo de desarrollo de resistencia del virus. Centrándonos en esto, la categoría terapéutica de los inhibidores antagonistas de los coreceptores CXCR4 y/o CCR5 es una buena diana de investigación. En la presente tesis se estudia la interacción de posibles candidatos a fármacos, inhibidores de los coreceptores CXCR4 y/o CCR5, los cuales eviten la unión de la glicoproteína gp120 del VIH a dichos coreceptores, impidiendo de esta forma la entrada del virus en las células.

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 coreceptores CXCR4 y CCR5, sus ligandos inhibidores antagonistas, y los complejos ligando-receptor, tratando de aportar información que ayude a entender las relaciones estructura-actividad. Para ello, se han llevado a cabo estudios de modelado de proteínas y la puesta a punto de una serie de filtros virtuales, basados tanto en el receptor como en el ligando, con los que cribar moléculas y seleccionar las posibles candidatos a fármacos.

Siguiendo el protocolo propuesto en la Introducción, se procede 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.

2.1. Modelos de los coreceptores CXCR4 y CCR5

Dado que no existen actualmente estructuras cristalográficas de los coreceptores CXCR4 y CCR5, se deben obtener modelos tridimensionales de dichas proteínas por homología a una proteína plantilla similar en estructura. Siguiendo la trayectoria de trabajos ya realizados en el laboratorio de diseño molecular del IQS [30] en los cuales se modelaron por homología los coreceptores CXCR4 y CCR5 tomando como plantilla la proteína rodopsina bovina perteneciente a la misma familia (GPCRs), se refinan estos modelos, concretamente los *loops*. Como resultado de la mejora de estos modelos previos se han obtenido:

Para CXCR4:

Modelo 1 CXCR4: Modelo por homología construido con Modeller [207] utilizando como plantilla la proteína rodopsina bovina (código PDB 1hzx). Los *loops* intermedios fueron construidos *ab initio* con Modeller [30] y refinados *ab initio* con Congen [212]. La metodología empleada se detalla en el Artículo I.

Modelo 2 CXCR4: *Modelo 1* refinado aplicando al complejo *modelo 1*- AMD3100 200 ps de dinámica molecular con Amber siguiendo el protocolo descrito en Orozco *et al.* [224] hasta equilibrado del sistema (Artículo I).

Para CCR5:

Modelo 1 CCR5: Modelo por homología construido con Modeller utilizando como plantilla la proteína rodopsina bovina (código PDB 1hzx). Los *loops* intermedios fueron construidos *ab initio* con Modeller [30] y refinados *ab initio* con Congen. La metodología empleada se detalla en el Artículo I.

Modelo 2 CCR5: *Modelo 1* refinado aplicando al complejo *modelo 1*- TAK779 200 ps de dinámica molecular con Amber siguiendo el protocolo descrito en Orozco *et al.* hasta equilibrado del sistema (Artículo I).

Modelo 3 CCR5: Modelo por homología construido con Moe-Align y Moe-Homology [260] utilizando como plantillas la proteína β2-adrenérgico (código PDB 2rh1) y la proteína rodopsina bovina (código PDB 1u19) para los *loops* intermedios. Dado que este modelo no mejora los resultados obtenidos por los anteriores en un cribado virtual retrospectivo, no se utiliza como filtro en el protocolo de cribado de potenciales inhibidores de entrada del VIH llevado a cabo en la presente tesis (Apéndice I).

En todos los casos, la estereoquímica ha sido comprobada con Procheck [215] mostrando que aproximadamente un 93,5% (CXCR4) y un 97,5% (CCR5) de los residuos cae dentro de las regiones favorables o permitidas del mapa de Ramachandran. El *modelo 1 CXCR4* muestra el 64% de los residuos en regiones favorables y el 29.3% en regiones permitidas. El *modelo 2 CXCR4* muestra el 64,2% de los residuos en regiones favorables y el 29,1% en permitidas. El *modelo 1 CCR5* muestra el 71,2% en favorables y el 26,3% en permitidas. El *modelo 2 CCR5* muestra el 70,4% en favorables y el 26,9% en permitidas. El *modelo 3 CCR5* muestra el 70,9% en favorables y el 26,8% en permitidas.

2.1.1. Consideraciones sobre los modelos obtenidos

Dado que un modelo nunca es correcto (p. ej. Modeller está diseñado para modelar proteínas globulares y no proteínas de membrana), se debe escoger el modelo que mejor interprete los datos que se desean. A continuación se discuten los modelos obtenidos y la metodología empleada.

Los programas como Modeller o Moe intentan encontrar *templates* con una identidad de secuencia alta, lo cual para los GPCRs es imposible, ya que solo existen dos plantillas (rodopsina bovina y β2-adrenérgico) cuya identidad de secuencia con la proteína diana está por debajo del 25%. Por lo tanto estos métodos de modelado aplicados a la región transmembrana tratando con GPCRs prácticamente sirven únicamente como un método de *threading* (superponer la secuencia diana sobre la estructura de la plantilla). Modeller normalmente hace una pequeña modificación estructural (generalmente cercana a los *loops*) basada en datos estructurales extraídos de proteínas globulares. Las conclusiones que se extraigan del modelo obtenido han de tener en cuenta este factor (sobretodo si las interfícies hélice-*loop* estan cercanas al sitio de unión y se utilizan explícitamente para estudios de *docking*) además de la influencia de las opciones escogidas en la optimización de geometría y minimización de energía.

Los modelos obtenidos son pues una "copia muy similar" de la rodopsina. Se podría describir como la secuencia de los coreceptores CXCR4 y CCR5 sobre la estructura de la rodopsina. Esto se considera correcto como aproximación ya que no se dispone de datos experimentales. Ahora bien,

hay que tener en cuenta que la estructura de la rodopsina ha evolucionado para traducir los cambios conformacionales del retinal a la región intracelular, y no para unir quimiocinas, con lo que los modelos obtenidos pueden estar muy lejanos de la realidad.

Por lo que respecta al modelado de los *loops*, es la parte más difícil, dado que la homología de secuencia es prácticamente inexistente en dichas regiones (excepto en ejemplos muy concretos entre familias de receptores específicas), por lo tanto los métodos de modelización por homología no son una buena opción. Por otro lado, cualquier método de modelización *ab intio* como Modeller o Congen (Figura 2.1) sólo es relativamente fiable si trata con *loops* cortos (<=4 aminoácidos), donde el espacio conformacional es particularmente pequeño. Para *loops* más largos estos métodos también son deficientes. Por lo tanto si se modelan *loops* cortos o *loops* que se encuentran lejanos a la zona de interés (el sitio de unión en el caso de este estudio), utilizando cualquier método se obtendrá un *loop* incorrecto pero que servirá para completar la estructura proteica sin poder hacer una predicción certera. Ahora bien, para modelar un *loop* largo (> 4 aminoácidos) y que forme parte del sitio de unión, como el *loop* extracelular II (E2) en las proteínas que se tratan en esta tesis, lo más correcto sería utilizar datos experimentales, como datos de accesibilidad de residuos (obtenidos por *cysteine scanning* por ejemplo) efectuados sobre los receptores CXCR4 y CCR5. Es posible incluir estos resultados en la confección del modelo con Modeller. Asimismo, si se conoce que ciertos residuos del *loop* son claves para la unión del ligando, se puede tener una idea de la orientación de dichos residuos en particular, lo cual puede incluirse también de manera explícita en el modelo utilizando Modeller (mediante un restricción de distancia por ejemplo). Así pues, cuanta más información experimental de los *loops* se disponga, mayor número de resultados experimentales será capaz de explicar el modelo, a pesar de que seguirá siendo solo un modelo.

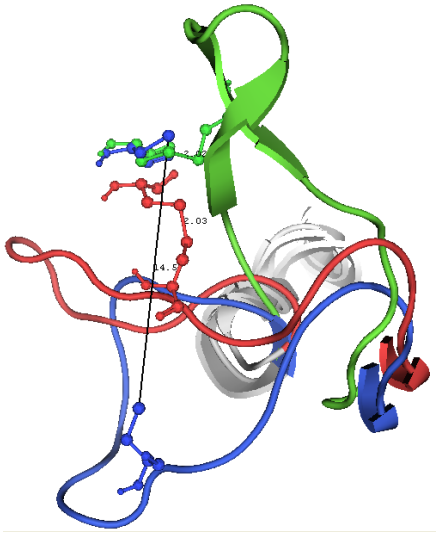


Figura 2.1 Detalle de la superposición del *loop* E2 del coreceptor CXCR4 modelado según: (1) el modelo obtenido *ab initio* con Modeller (verde); los residuos cisteína se muestran en representación de *balls & sticks* en verde; los azufres que forman el puente disulfuro se encuentran a distancia de 2.02 Å. (2) El modelo obtenido *ab initio* con Congen (azul) sin tener en cuenta restricciones de puente disulfuro (los residuos cisteína se muestran en representación de *balls & sticks* en azul; los azufres que forman el puente disulfuro se encuentran a distancia de 14.57 Å. (3) El modelo obtenido *ab initio* con Congen (rojo) teniendo en cuenta restricciones de puente disulfuro; los residuos cisteína se muestran en representación de *balls & sticks* en rojo; los azufres que forman el puente disulfuro se encuentran a distancia de 2.03 Å. Se puede observar como el análisis conformacional sistemático que ejecuta Congen permite obtener una conformación del *loop* E2 más abierta, la cual no cause impedimento estérico a la entrada de ligandos a la cavidad de unión al aplicar técnicas de *docking*.

Por ejemplo, comparando las estructuras de la rodopsina y el β2-adrenérgico (Figura 2.2), se observa que el E2 es muy diferente entre ambas. Hubiera sido imposible predecir la estructura del E2 del β2-adrenérgico a partir del E2 de la rodopsina. De la misma manera, muy probablemente es extraordinariamente difícil predecir la estructura del E2 de un receptor de quimiocinas a partir del E2 de la rodopsina o del β2-adrenérgico. A pesar de que existen características comunes (el puente de cisteína entre el E2 y la hélice III, o, entre rodopsina y el β2-adrenérgico, la longitud del *loop*), este *loop* es clave en el proceso de reconocimiento del ligando. En la rodopsina dicho *loop* forma parte del sitio de unión y, principalmente, bloquea al retinal impidiendo su salida. En cambio, en el β2-adrenérgico el *loop* E2 deja espacio para que el sitio de unión sea accesible a los ligandos extracelulares. En lo referente a los receptores de quimiocinas parece lógico pensar que el *loop* E2 también deba dejar espacio para que entre el extremo *N*-terminal de las quimiocinas. Además se podría suponer que muy probablemente tenga que tener una estructura determinada que favorezca la unión del resto de la quimiocina a la región extracelular del receptor.

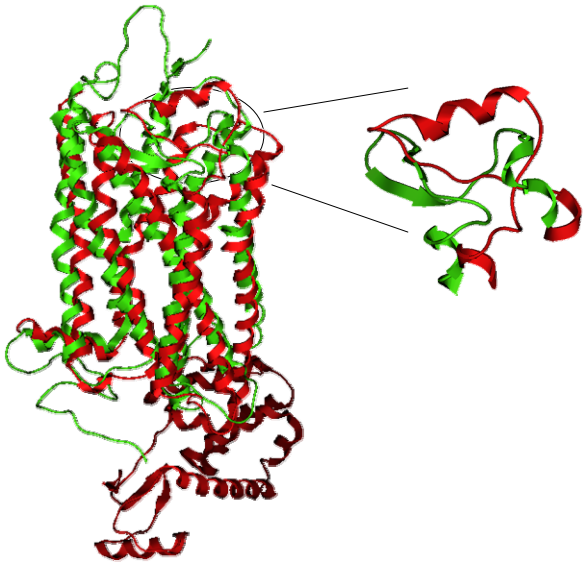


Figura 2.2 Comparación del *loop* extracelular 2 (E2) de las estructuras de la rodopsina bovina (código PDB 1u19), representada en verde, y el β2-adrenérgico (código PDB 2rh1), representado en rojo.

Así pues, cabe prestar mayor atención al estudio del sitio de unión, el cual es el que va a tener repercusión en el cribado virtual. Las otras regiones, como los *loops* o extremos *N* y *C*-terminales no sería necesario modelarlas si no intervienen en el sitio activo, o bien se pueden modelar, de tal manera que no influyan en la estructura de las zonas que interesan, con el fin de tener un modelo de la secuencia completa. Es importante tener en cuenta que si se modelan las regiones que no intervienen en el sitio activo se haga de manera que no distorsionen los resultados del cribado virtual (por ejemplo mostrando interacciones de algunos residuos con algún tipo de ligandos que realmente no exista), dado que los resultados que se van a obtener pueden estar lejanos a la realidad.

Centrándonos en el caso concreto de este trabajo, es especialmente importante tener en cuenta la presencia de interacciones entre residuos altamente conservados, p. ej. N1.50, D2.50 i N7.49 (en el sistema general de numeración de Ballesteros [187]), residuos que están presentes en virtualmente todos los GPCRs, y que están involucrados en una red de interacciones mediada por moléculas de agua (tanto en la rodopsina como en el β2-adrenérgico). Además se han de mantener ciertas

distorsiones estructurales dentro de las hélices, las cuales no se pueden predecir *ab-initio*, y que, al involucrar también residuos altamente conservados, es de esperar que sean comunes a muchas familias de GPCRs. Por ejemplo, las distorsiones debidas a las prolinas altamente conservadas de las hélices V, VI y VII provocan distorsiones que no tienen nada que ver con un *proline-kink* "estándar". Además de cierta torsión provocan fuertes distorsiones locales (oberturas en las hélices V y VI, y cerramiento en la hélice VII), en algunos casos estabilizadas por moléculas de agua (véase Figura 2.3) [272].

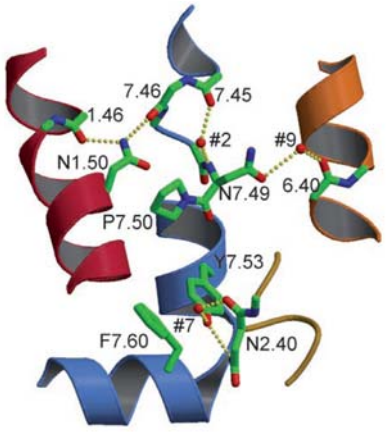


Figura 2.3 Estructura cristalina de la rodopsina (PDB ID: 1GZM). Se muestran las transmembranas 1 (rojo), 2 (amarillo), 6 (naranja), 7 (azul), y la hélice α contenida en el extremo *C*-terminal (azul) de la rodopsina. Se observan las interacciones que implican a los residuos aminoácidos altamente conservados de dichas transmembranas y la hélice α del extremo *C*-terminal. Asimismo se muestra la participación de moléculas de agua, etiquetadas como #, en la estabilización de dichas interacciones. Extraído de [272].

Así pues, como primera aproximación parece buena idea utilizar la estructura de la rodopsina como se ha procedido en esta tesis, manteniendo las interacciones inter-helicales clave respetando las conformaciones de los residuos conservados respecto a la rodopsina y las distorsiones alrededor de ellos, ya que de esta manera se conservan determinadas especificidades estructurales que son claves en todos los GPCRs. Ahora bien, se podría mejorar el modelo no solo incluyendo las similitudes tal como se ha hecho sino también las diferencias, cosa que hace que un receptor de quimiocinas una quimiocinas y no aminas biogénicas. Con el objeto de tratar las diferencias entre el receptor objetivo y el patrón de rodopsina se puede partir del patrón buscando motivos de secuencia específicos que sean diferentes entre las dos familias (rodopsinas y receptores de quimiocinas en el caso del presente estudio) e intentar encontrar cual es su papel estructural con soporte experimental [273, 274].

En cuanto al patrón utilizado, cabe decir que la estructura publicada del β2-adrenérgico tiene unido un agonista inverso, el carazolol, el cual está muy bien empaquetado en el sitio de unión (Figura 2.4). Este ligando es capaz de dar mucha estabilidad al receptor por la cual cosa se han podido obtener los cristales. Ahora bien, la estructura del sitio de unión es muy probablemente característica del modo de unión de este ligando, que no es uno de los ligandos naturales del β2-adrenérgico. Así pues, en mayor o menor medida, el sitio de unión del β2-adrenérgico está distorsionado respecto al receptor *wild-type*. Por lo tanto, utilizar esta estructura tal cual para modelar otro receptor comporta introducir todos estos cambios estructurales que muy posiblemente son únicos del β2-adrenérgico unido al carazolol. Ello lleva al concepto de *induced-fit* [36] (reorganización del receptor debido a la unión del ligando), lo cual constituye un factor complicado en el cribado virtual basado en la estructura, dado que una buena predicción de este fenómeno

requiere tener en consideración la flexibilidad del receptor, la cual se tiene en cuenta solo parcialmente en la mayoría de programas de *docking*. Por lo que respecta a la rodopsina como patrón, es una plantilla que une a otro agonista inverso, el 9-*cis*-retinal a pesar de que en este caso se trata de un agonista inverso natural. En definitiva, utilizando una u otra plantilla, en algunas regiones será semejante al receptor diana (alrededor de los residuos altamente conservados y hacia el lado citoplasmático del *transmembrane bundle*) y en algunas otras será diferente (alrededor del sitio de unión) [275]. Por lo tanto hay que ser muy consciente de lo que se está haciendo, las limitaciones que se están considerando, las similitudes y diferencias que se están aceptando y la decisión de cómo tratarlas, siempre valorando los resultados que se obtienen de los algoritmos automáticos de modelización.

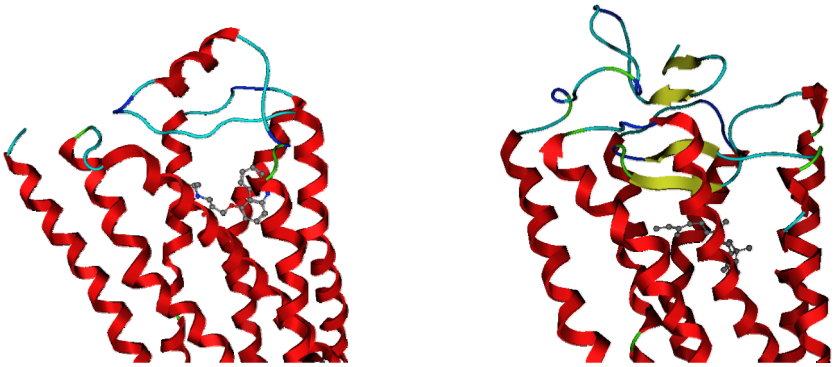


Figura 2.4 Carazolol en la cavidad de unión del β2-adrenérgico (código PDB 2rh1), izquierda. Retinal en la cavidad de unión de la rodopsina bovina (código PDB 1u19), derecha.

Teniendo en cuenta los argumentos expuestos anteriormente, el *modelo 1 CXCR4* y el *modelo 1 CCR5* (Figura 2.5) son los que han sido utilizados para el posterior cribado virtual basado en la estructura.

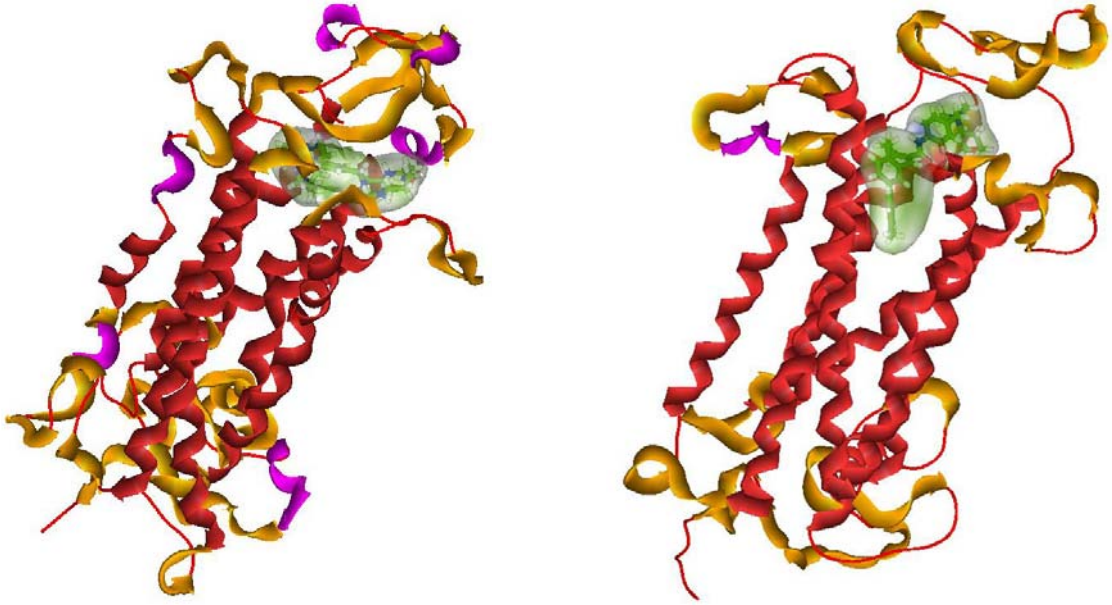


Figura 2.5 *Modelo 1 CXCR4* (izquierda) y *modelo 1 CCR5* (derecha) con los ligandos AMD3100 y TAK779, respectivamente, situados en la cavidad de unión mediante *docking*.

2.1.2. Estudios de *docking* y dinámica molecular

Con el objeto de validar los modelos obtenidos de los coreceptores CXCR4 y CCR5 se ha realizado un estudio de *docking* de dichos coreceptores con inhibidores de entrada ya conocidos (AMD3100 para CXCR4 y TAK779 para CCR5). Es razonable asumir que si los modelos construidos son relativamente correctos un protocolo automatizado de *docking* debe de ser capaz de ubicar de manera correcta el sitio de unión y relativamente correcta el modo de unión de un ligando activo. Por ello se estudia el sitio de unión de CXCR4 y CCR5 mediante *docking* ciego con Autodock y el modo de unión mediante *docking* tradicional con Autodock y Gold. Los resultados obtenidos, así como el protocolo utilizado se detallan en el Artículo I. Cabe mencionar que el sitio de unión obtenido para ambos coreceptores está perfectamente de acuerdo con los estudios de mutagénesis dirigida existentes [135, 162-176] (véase Sección I.2.6 e I.2.7). Asimismo, el modo de unión obtenido para CXCR4 y CCR5 se asemeja a estudios computacionales realizados previamente por otros grupos investigadores [168-177]. El Apéndice II muestra un esquema gráfico de estos resultados presentado como comunicación oral en el IX Congress of the ISJACHEM.

A fin de refinar el modo de unión obtenido para AMD3100 y TAK779, se ha realizado una dinámica molecular utilizando Amber 8 a partir de las mejores poses de *docking* CXCR4-AMD3100 y CCR5-TAK779 siguiendo el protocolo descrito en Orozco *et al* [224]. Se han obtenido poses con una distancia promedio de 2 Å más cercana a los residuos claves para la unión de dichos ligandos según los estudios de mutagénesis dirigida. La Figura 2.6 muestra los resultados así como los gráficos de equilibrado del sistema, clave para obtener una buena calidad de la trayectoria.

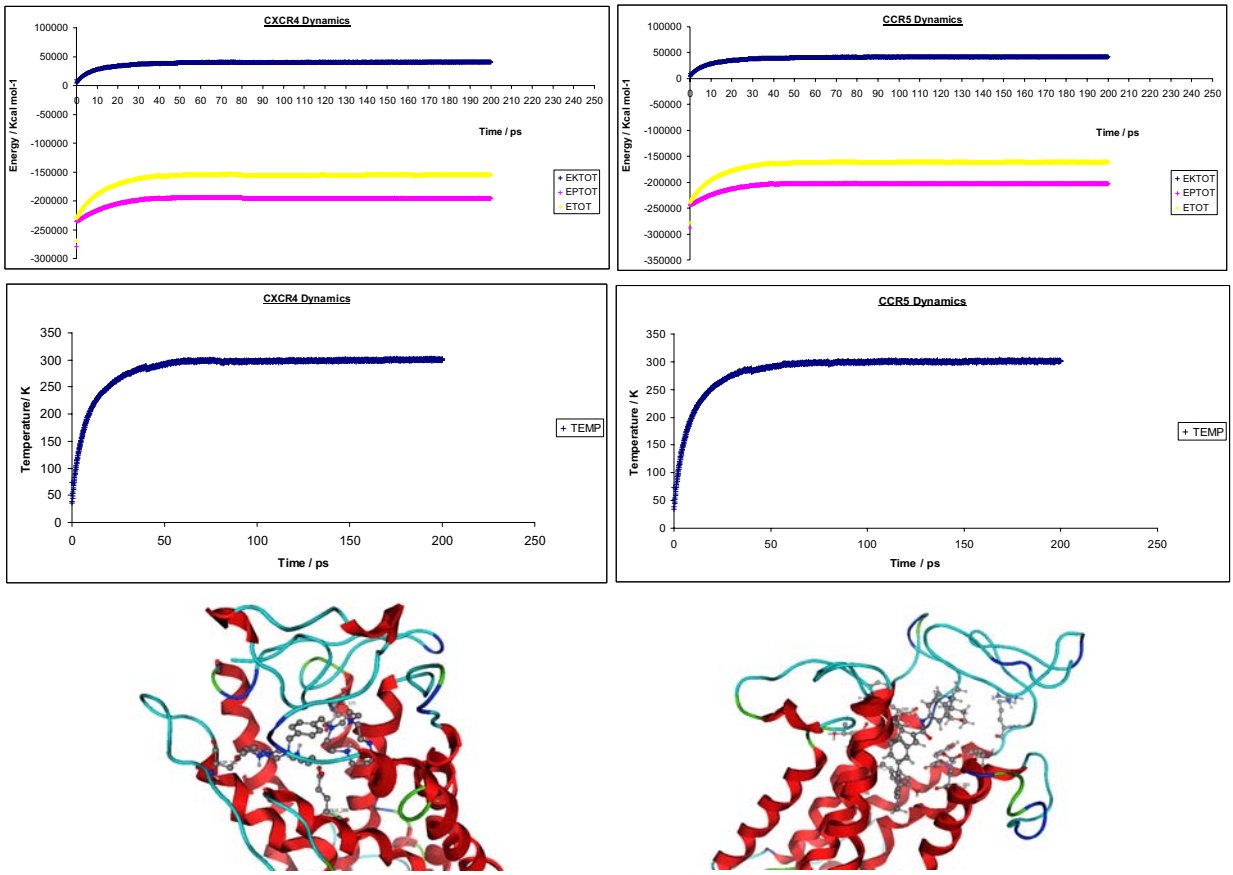


Figura 2.6 *Modelo 2 CXCR4* (izquierda) y *modelo 2 CCR5* (derecha) con los ligandos AMD3100 y TAK779, respectivamente, situados en el sitio de unión mediante *docking* y refinado por dinámica molecular.

Dado que los estudios de *docking* y refinado por dinámica molecular del sitio y modo de unión de los coreceptores CXCR4 y CCR5 son compatibles con los datos de mutagénesis dirigida existentes, se puede considerar en términos generales que los modelos construidos podrían ser inicialmente aplicados en un cribado virtual con el fin de detectar compuestos antagonistas de CXCR4 y CCR5.

2.2. Cribado Virtual Retrospectivo

Con el fin de poder distinguir moléculas activas e inactivas frente a la inhibición de los coreceptores de estudio, y por lo tanto de la entrada del VIH a las células, se realiza un cribado virtual de compuestos. En primer término, de manera retrospectiva, es decir, cribando virtualmente una biblioteca previamente compilada de inhibidores conocidos de CXCR4 y CCR5 y compuestos tipo fármaco supuestamente inactivos. Para cada receptor se aplican técnicas de cribado virtual *receptor-bassed* (*docking*) y *ligand-based* (*shape matching*, modelos farmacofóricos, y búsquedas de similitud) utilizando ligandos activos como referencia, y se analizan los factores de enriquecimiento y diversidad de las listas de *hits* resultantes.

Se utiliza una base de datos de 602 compuestos inhibidores antagonistas de CXCR4 y CCR5 compilada a partir de bibliografía y 4696 compuestos tipo fármaco con propiedades 1D similares a las de los activos (masa molecular, número de dadores de hidrógeno, número de aceptores de hidrógeno, número de enlaces simples rotables, número de átomos hidrofóbicos, coeficiente de partición octanol-agua) extraídos de *Maybridge Screening Collection* [276]. Las características de esta base de datos, así como los resultados del cribado que con ella se obtienen se encuentran en el Artículo I. Asimismo se aplican las mismas técnicas de cribado virtual a dos familias concretas de inhibidores de CCR5 pertenecientes a la base de datos *Berlex* (derivados de 4-hidroxipiperidina [277] y guanilhidrazonas [278]) cribadas junto a 3388 compuestos tipo fármaco también extraídos de *Maybridge* con propiedades 1D similares a las de los activos de *Berlex*. El Artículo V muestra los resultados obtenidos en este caso. Más adelante se amplia la base de datos inicial a 672 antagonistas de CXCR4 y CCR5 y se aplican las anteriores aproximaciones y una nueva técnica de cribado virtual *shape matching* (véase Sección 1.7) diseñada en esta tesis. El Artículo II expone los resultados obtenidos así como la nueva técnica de *shape matching* implementada.

Los resultados obtenidos muestran que, en general, las aproximaciones *ligand-based shape matching* utilizadas (Parafit, Rocs, Hex) obtienen mayores enriquecimientos que las *receptor-based docking* (Autodock, Gold, Hex, Fred), especialmente para CXCR4, dada la dificultad de encontrar buenas conformaciones para los inhibidores de CCR5 (véase Figura 2.7). Los resultados obtenidos para CCR5 sugieren la posibilidad de que diferentes *scaffolds* activos se unan de diferentes maneras al sitio de unión de CCR5. Estudios computacionales previos parecen corroborar la existencia de más de un sitio de unión dentro de la cavidad de unión de CCR5 [135, 172, 173, 174, 177, 182-186] (véase Sección I.2.7). Esta hipótesis se estudia de manera detallada en el Artículo II mediante el uso de la nueva aproximación de *clustering* basada en la obtención de una forma esférica harmónica promedio (*spherical harmonic-based consensus shape clustering approach*) diseñada en esta tesis (véase Sección 2.2.2). Siguiendo esta idea se construyen múltiples combinaciones de ligandos inhibidores de CCR5 en múltiples superposiciones con el fin de encontrar una *query* promedio que sea capaz de obtener elevados enriquecimientos en el cribado virtual (véase Figura 2.8).

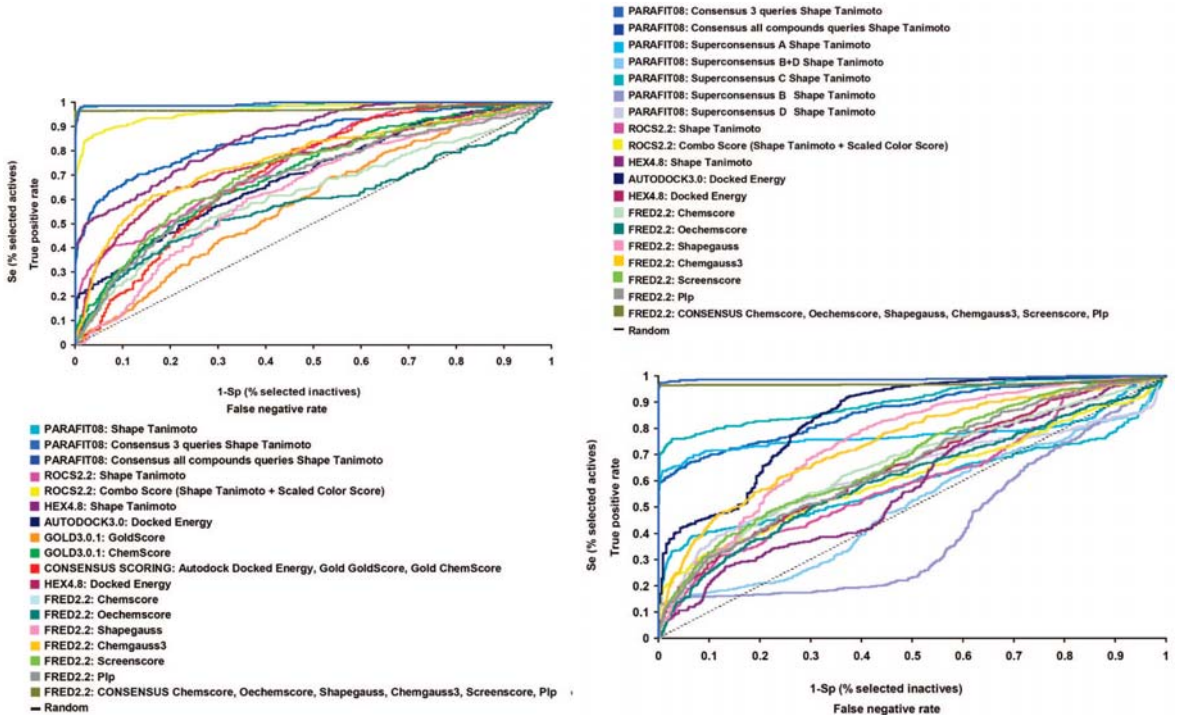


Figura 2.7 Gráficos ROC (sensibilidad o tasa de positivos reales versus 1- especificidad o tasa de falsos positivos) de los resultados obtenidos para el cribado virtual retrospectivo de 672 inhibidores de los co-receptores CXCR4 (izquierda) y CCR5 (derecha) utilizando técnicas *structure-based docking* y *ligand-based shape matching*.

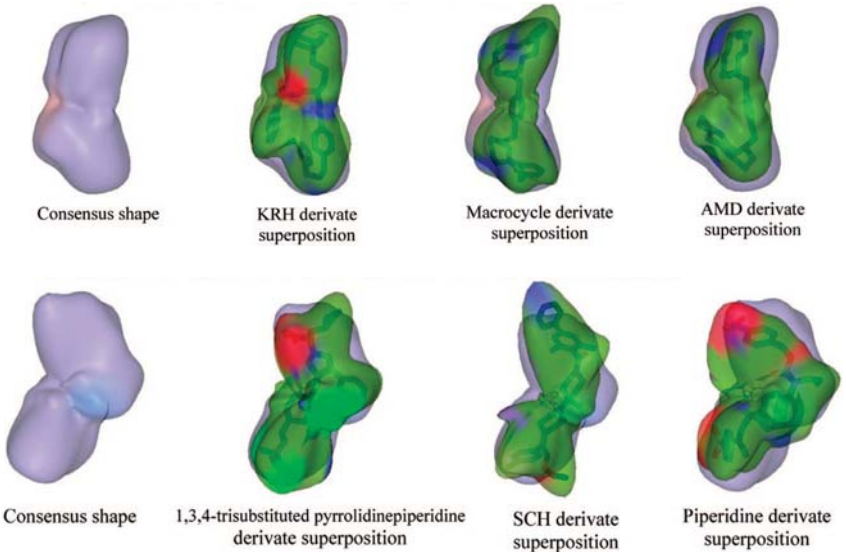


Figura 2.8 Nueva aproximación de *clustering* basada en la obtención de una forma esférica harmónica promedio diseñada en esta tesis. La imagen de la izquierda muestra la forma promedio obtenida a partir del consenso de formas de los tres compuestos más activos de la base de datos de inhibidores de CXCR4 (arriba) y CCR5 (abajo). Las siguientes tres imágenes muestran la superposición de estos tres compuestos sobre el consenso, para ambos casos.

Se procede a discutir en las siguientes secciones los resultados obtenidos en el cribado virtual retrospectivo según las diferentes aproximaciones utilizadas. El Apéndice IV (póster presentado en el 4th Joint Sheffield Conference on Chemoinformatics) y el Apéndice V muestran un resumen gráfico de dichos resultados.

2.2.1. *Structure-Based*

Tras haber estudiado bien el sitio y el modo de unión de los inhibidores de los coreceptores CXCR4 y CCR5 y tener bien caracterizado el *docking* con dichos coreceptores, de manera que reproduzca modos de unión fiables proteína-ligando, se puede proceder a realizar un cribado virtual *structure-based* de compuestos.

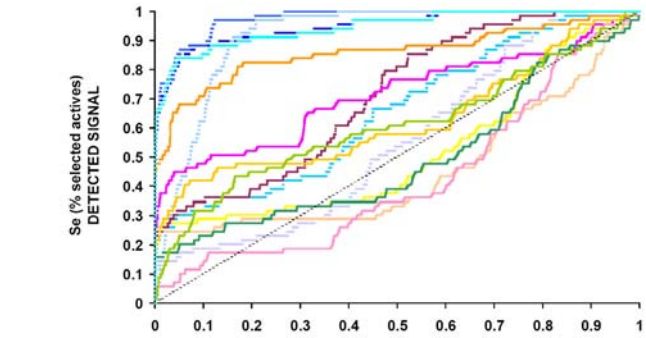
De este modo se analiza la habilidad de los modelos construidos para discriminar compuestos activos de inactivos mediante el cálculo de curvas de enriquecimiento y gráficos ROC (*receiver-operator-characteristic plot*) utilizando las funciones de *scoring* Autodock Docked Energy y Fred Plp, Chemgauss3, Shapegauss, OechemScore, ScreenScore, ChemScore, y Consensus scoring para CXCR4 y CCR5. Además, las curvas de enriquecimiento y gráficos ROC para CXCR4 han sido también calculados utilizando las funciones de *scoring* Gold GoldScore, ChemScore, y Rank-by-Rank Consensus scoring (véase Figura 2.7). El Artículo I muestra las curvas de enriquecimiento obtenidas utilizando dichas funciones de *scoring* y la discusión detallada para cada una de ellas. En global, los resultados de este estudio indican que los modelos de CXCR4 y CCR5 y el supuesto modo de unión para las moléculas activas se pueden considerar útiles dado que las curvas de enriquecimiento muestran un buen reconocimiento de los activos conocidos. Ahora bien, las curvas de enriquecimiento para CCR5 no son igual de buenas en algunos casos. El coreceptor CCR5 parece tener un mayor *binding pocket* que CXCR4 y por ello es difícil para los algoritmos de *docking* ubicar modos de unión factibles para los activos conocidos. Además, análisis de dinámica molecular del E2 de CCR5 muestran que esta región es altamente flexible y puede servir como una tapadera flexible que restrinja a los ligandos dentro de la cavidad de unión adyacente. Así pues, la calidad de los modelos de los receptores, especialmente en las regiones modeladas correspondientes a los *loops*, es crítica para el reconocimiento de ligandos conocidos. Los resultados muestran que si se puede construir un modelo relativamente correcto de receptor, como en el caso de CXCR4, se pueden obtener resultados de enriquecimiento respetables comparados con otros estudios de cribado virtual basado en *docking* los cuales utilizan GPCRs modelados. Por otro lado, las proteínas modeladas por homología inevitablemente contendrán errores o imprecisiones. Por ello es necesario complementar los estudios *structure-based docking* con técnicas de cribado virtual *ligand-based* si se dispone de un conjunto de activos conocidos. Los resultados, basados en datos de ligandos activos conocidos, obtenidos por dichas técnicas deberán contraponerse con los resultados *structure-based* y los problemas de modelado. La siguiente sección se ocupa de este punto.

2.2.2. *Ligand-Based*

Las estrategias utilizadas en el cribado virtual *ligand-based* comprenden búsquedas de similitud respecto a ligandos activos conocidos, desarrollo de modelos farmacofóricos y *shape matching*. La implementación de esta última ha sido motivo de una estancia predoctoral en la Universidad de Aberdeen, colaborando con el grupo del Dr. Dave Ritchie, creador de los softwares Parafit y Hex [267], los cuales utilizan esta técnica (véase Sección 1.7). Se procede a discutir los resultados obtenidos con cada una de estas aproximaciones.

Las búsquedas de similitud se han llevado a cabo con Qikprop/Qiksim de la empresa Schrödinger [261, 279], a través del Servei de Disseny de Fàrmacs (SDF) accediendo remotamente al Centre de Supercomputació de Catalunya (CESCA). Los parámetros utilizados, así como los resultados obtenidos expresados en un gráfico ROC (*receiver-operator-characteristic plot*) se detallan en el Artículo V. Cabe destacar que esta aproximación de cribado virtual rinde resultados comparables, aunque aún inferiores, a las técnicas de *shape matching* utilizadas, cuando se utiliza como *query* el

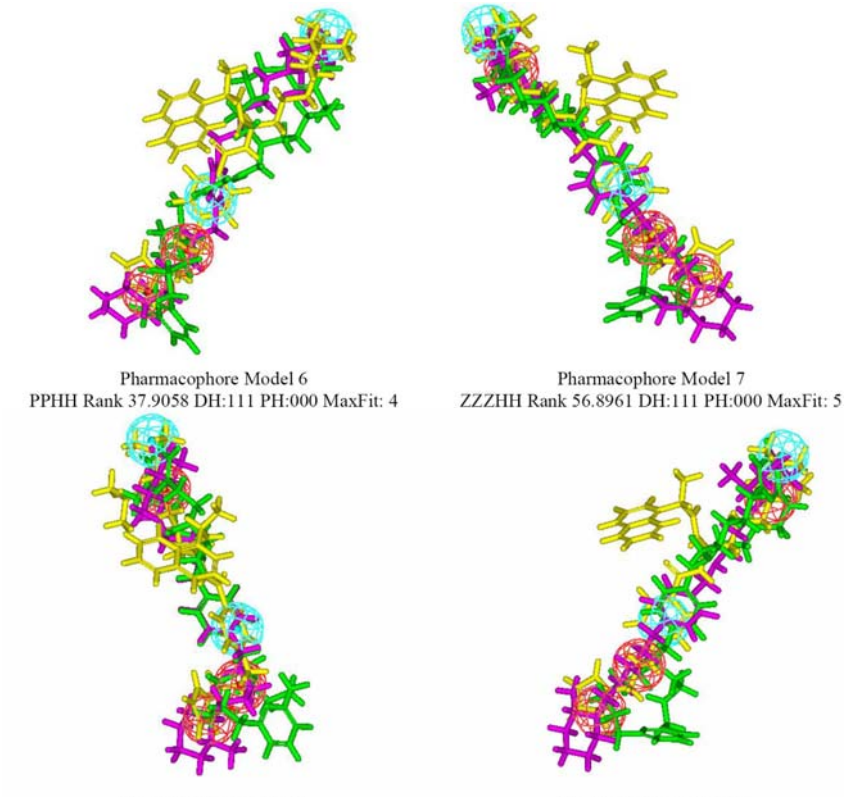
promedio de las propiedades fisicoquímicas y biológicas calculadas para todos los activos (véase Figura 2.9)



Scoring Function	AUC
PARAFIT 8.02v: Shape consensus Tanimoto (3 active compounds)	0.976
PARAFIT 8.02v: Shape consensus Tanimoto (all active compounds)	0.951
PARAFIT 8.02v: Shape consensus Tanimoto (6 PCA selected compounds)	0.947
PARAFIT 8.02v: Shape Tanimoto (query cmpd10h docked onto CCR5 model)	0.918
ROCS2.2: Combo Score (Shape Tanimoto + Scaled Color Score) 10 query conf & 10 conf/database compound	0.861
ROCS2.2: Shape Tanimoto 10 query conf & 10 conf/database compound(query cmpd10h from 3D-QSAR)	0.714
HEX4.8: Shape Tanimoto (query cmpd10h from 3D-QSAR)	0.710
PARAFIT 8.02v: Shape Tanimoto (query cmpd10h from 3D-QSAR)	0.655
ROCS2.2: Combo Score (Shape Tanimoto + Scaled Color Score) 10 conf/database compound	0.628
QIKPROP/QIKSIM (actives averaged descriptors): Tanimoto Score	0.609
PARAFIT 8.02v: Shape Tanimoto (query TAK779 docked onto CCR5 model)	0.555
ROCS2.2: Combo Score (Shape Tanimoto + Scaled Color Score) (query cmpd10h from 3D-QSAR)	0.521
QIKPROP/QIKSIM (cmpd10h descriptors): Tanimoto Score	0.495
ROCS2.2: Shape Tanimoto (query cmpd10h from 3D-QSAR)	0.456
ROCS2.2: Shape Tanimoto 10 conf/database compound(query cmpd10h from 3D-QSAR)	0.426

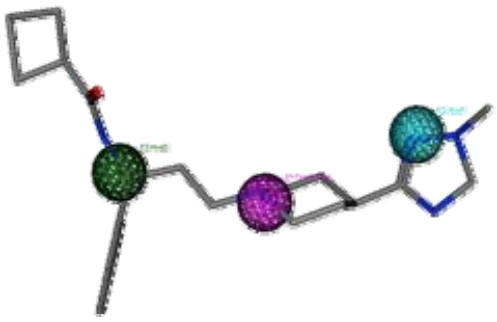
Figura 2.9 Gráfico ROC (sensibilidad o tasa de positivos reales versus 1- especificidad o tasa de falsos positivos) de los resultados obtenidos para el cribado virtual retrospectivo de 69 inhibidores derivados de guanihidrazona y 4-hydroxipiperidina del coreceptor CCR5 utilizando las técnicas *structure-based docking* y *ligand-based shape matching* y búsquedas de similitud. AUC se refiere al área hallada bajo la curva.

Por lo que respecta al desarrollo de modelos farmacofóricos como filtro para cribar virtualmente, se han construido diversos farmacóforos utilizando Moe (para CCR5) y Discovery Studio (para CXCR4) [264], este último accesible también a través del SDF con conexión remota al CESCA (véanse Figuras 2.10, y 2.11 respectivamente). El Artículo III detalla la metodología aplicada para la construcción de los modelos farmacofóricos para CXCR4, así como las características de cada uno. Cabe mencionar que este artículo también presenta modelos farmacofóricos de CXCR4 realizados con Moe (modelos 1 a 5), los cuales han sido desarrollados por S. Pettersson, coautora del artículo. La misma metodología se ha empleado para desarrollar modelos farmacofóricos con Moe para el coreceptor CCR5. El Artículo V muestra estos resultados. A fin de validar retrospectivamente los modelos farmacofóricos para CXCR4 y CCR5, se han calculado una serie de parámetros detallados también en los Artículo III y V. Estos parámetros comprenden: *A%,* el porcentaje de activos encontrados en la lista de *hits* (H_a) respecto al total de activos (A); *Y%,* porcentaje de activos encontrados en la lista de *hits* respecto al total de *hits* encontrados (H_t); enriquecimiento (*EF*); bondad de la lista de *hits (GH)*; falsos negativos (*false* -); falsos positivos (*false* +). La elección del modelo utilizado para cribar más adelante prospectivamente una base de datos se realiza observando principalmente los falsos positivos y falsos negativos que se obtienen del análisis retrospectivo. Se considera mejor modelo cuanto menor sea el número de falsos positivos y falsos negativos. En resumen, los mejores modelos para CXCR4 y CCR5 considerados clasifican bastante bien los activos conocidos. La inspección visual de la lista de *hits* activos resultantes del análisis retrospectivo muestra que el *ranking* de cada compuesto depende del modelo farmacofórico utilizado y la familia química a la que pertenece.



Model	H_a	H_t	false +	false -	*EF*	*Y(%)*	*A(%)*	*GH*
6	132	132	0	19	10.68	100	87	0.97
7	96	96	0	55	10.68	100	64	0.91
8	106	107	1	45	10.58	99	70	0.92
9	92	92	0	59	10.68	100	61	0.90

Figura 2.10 Modelos farmacofóricos obtenidos con Discovery Studio para CXCR4. Muestran *features* (características) hidrofóbicas (H) en azul, cargadas positivamente (P) e ionizables positivamente (Z) en rojo, y anillos aromáticos (R) en naranja. Los compuestos activos frente a CXCR4 utilizados para construir las *queries* están superpuestos y alineados sobre la *query* farmacofórica (derivados de azamacrociclo en verde, KRH en amarillo, y diamina en rosa). Se muestra el resultado de la hipótesis calculado con HipHop para cada uno de los modelos y la tabla de parámetros calculados para validar los modelos.



Model	1	2	3	4	5	6
A[a]	332	332	332	332	332	332
D[b]	3650	3650	3650	3650	3650	3650
H_a[c]	243	231	173	273	275	261
H_t[d]	279	256	192	440	501	414
FP [e]	36	25	19	167	226	153
FN [f]	89	101	159	59	57	71
EF[g]	9.58	9.92	9.91	6.82	6.03	6.93
Y(%)[h]	87	90	90	62	55	63
A(%)[i]	73	70	52	82	83	79
GH[j]	0.83	0.84	0.80	0.64	0.58	0.64

Figura 2.11 Mejor modelo farmacofórico obtenido con Moe para el coreceptor CCR5 (*modelo 1*). El modelo muestra una *feature* o característica hidrofóbica (Hyd) en verde, un dador de puente de hidrógeno (Don) en lila, y un aceptor de puente de hidrógeno (Acc) en azul. Se muestra la superposición del *hit* más activo sobre la *query* farmacofórica. Se muestra también la tabla de parámetros calculados para validar los modelos construidos.

El *shape matching* se ha llevado a cabo utilizando Parafit, Rocs y Hex mediante la superposición de cada uno de los compuestos de la base de datos frente a una molécula *query* dada (las conformaciones obtenidas del *docking* de TAK779 y AMD3100 para CCR5 y CXCR4, respectivamente), véase Figura 2.12. La metodología empleada, así como la discusión de los resultados obtenidos para cada una de las aproximaciones se detallan en el Artículo I. Asimismo, se han aplicado las mismas aproximaciones al caso particular de derivados de 4-hidroxipiperidina y guanilhidrazonas, inhibidores conocidos de CCR5. La metodología empleada y los resultados obtenidos en este caso se muestran en el Artículo V. El Apéndice VI presentado como comunicación oral en XXth International Symposium on Medicinal Chemistry EFMC-ISMC resume los resultados de manera gráfica.

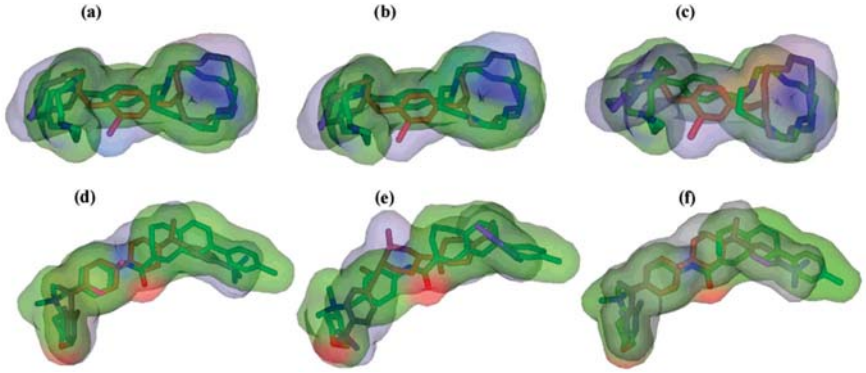


Figura 2.12 Ejemplo de superposiciones calculadas para el cribado virtual *shape matching*. Cada compuesto de la base de datos se muestra en azul/rojo, y el color de la molécula *quey* depende del tipo de átomo (azul, nitrógeno; verde, carbono; rojo, oxigeno). Las imágenes a, b, y c muestran un compuesto de la base de datos (AMD3167) superpuesto al AMD3100 con Parafit, Rocs, y Hex, respectivamente; las imágenes d, e, y f muestran un compuesto de la base de datos (derivado de SCH417690) superpuesto a TAK779 usando el mismo software, respectivamente.

En la mayoría de aproximaciones *shape matching* utilizadas para cribar virtualmente bases de datos, la forma cristalográfica del ligando extraída del complejo es la que se utiliza a menudo como molécula *query*. En este estudio, dado que no se dispone de información cristalográfica de las dianas, la molécula *query* se ha seleccionado a partir de la mejor conformación de unión obtenida del *docking* compatible con los estudios de mutagénesis dirigida, tal y como se ha mencionado anteriormente. Siguiendo con esta idea, es importante destacar el problema generalizado que comporta la selección de una conformación de la *query* apropiada para llevar a cabo un cribado virtual *shape matching* [280, 281]. En este aspecto, los resultados obtenidos muestran que es difícil obtener superposiciones satisfactorias de todos los antagonistas conocidos de CCR5 presentes en la base de datos compilada. Ninguna de las múltiples conformaciones de la *query* TAK779 utilizadas se superpone bien a todos los activos conocidos. No obstante, se conocen múltiples compuestos activos para CCR5 y todos ellos deben de encajar físicamente en la misma cavidad de unión. Por otra parte, los residuos claves según los estudios de mutagénesis dirigida están bien distribuidos espacialmente alrededor de la cavidad de unión de CCR5, lo cual sugiere una gran región de unión para los ligandos. Con tal de unificar estas observaciones se ha propuesto la hipótesis de que puede existir más de una región de unión dentro de la cavidad de CCR5, lo cual puede hacer que los ligandos activos se distribuyan de diferente manera (diversos *clusters*) dentro de la cavidad. Ello explicaría la dificultad de encontrar una superposición global satisfactoria de todos los inhibidores antagonistas conocidos de CCR5. Esta hipótesis de la multi-región de unión está también soportada por otros grupos de investigación [182, 184] (véase Sección I.2.7). Dicha hipótesis se ha estudiado mediante el desarrollo de la nueva técnica *spherical harmonic-based consensus shape matching*

mencionada anteriormente (Sección 2.2), la cual es capaz de identificar *clusters* de *scaffolds* de entre un gran conjunto de activos conocidos. El resultado de esta técnica aplicada a la base de datos compilada en esta tesis de 424 inhibidores conocidos de CCR5 rinde cuatro *shape consensus clusters* (véase Figura 2.13). Estudios de *blind docking* de estas cuatro "pseudomoléculas", generadas como la forma promedio de los compuestos más similares de la base de datos, predicen que se unen a tres subsitios solapados diferentes dentro de la cavidad de unión de CCR5 (véase Figura 2.14). Por lo tanto, las estructuras consenso (*super consensus*) calculadas proporcionan un fuerte apoyo a la hipótesis de la multi-región de unión del CCR5, y ayudan a tener una mejor idea de cómo probablemente se distribuyen los antagonistas de CCR5 en la cavidad de unión. El Artículo II expone detalladamente esta hipótesis, la aproximación *consensus shape* implementada y su aplicación a los inhibidores de CXCR4 y CCR5. Los resultados obtenidos muestran que una *query* promedio que sea capaz de combinar las características más significativas en cuanto a la forma de un conjunto de moléculas rinde resultados que mejoran significativamente los obtenidos utilizando una *query* convencional, sobretodo en dianas como el coreceptor CCR5, las cuales poseen grandes cavidades de unión (véase Figura 2.7). El Apéndice VII (póster presentado en Targeting and tinkering with Interaction Networks. From interaction networks to therapeutics) muestra un resumen gráfico de la nueva aproximación implementada en esta tesis.

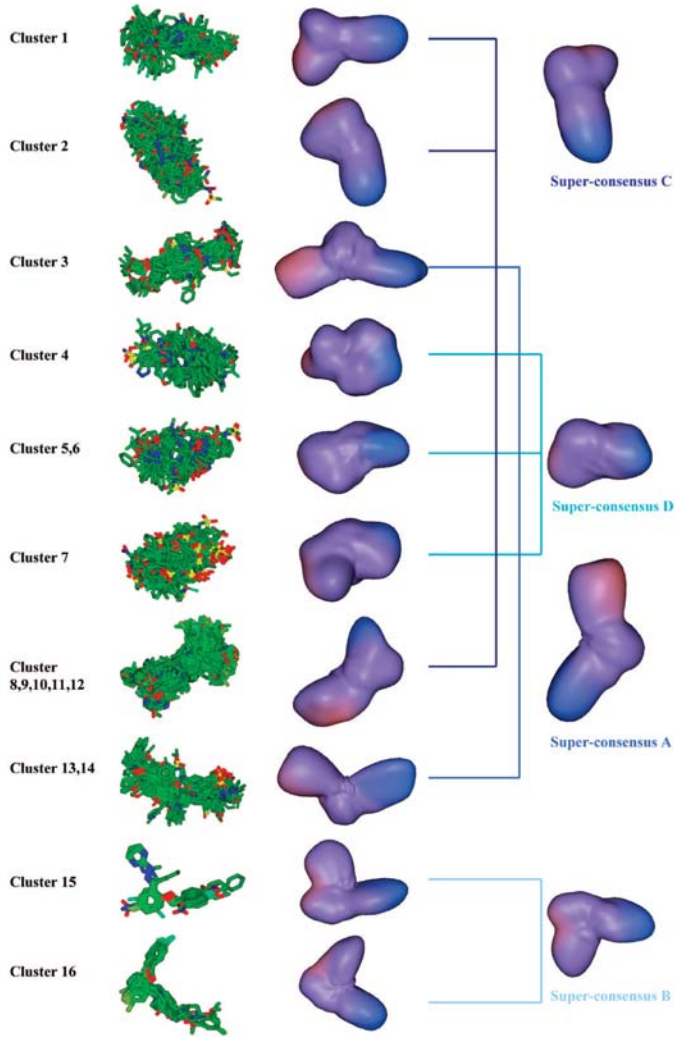


Figura 2.13 Superposiciones moleculares, formas consenso de los *clusters* identificados y "pseudomoléculas" consenso finales creadas a partir de 424 inhibidores conocidos de CCR5 pertenecientes a 15 familias diferentes.

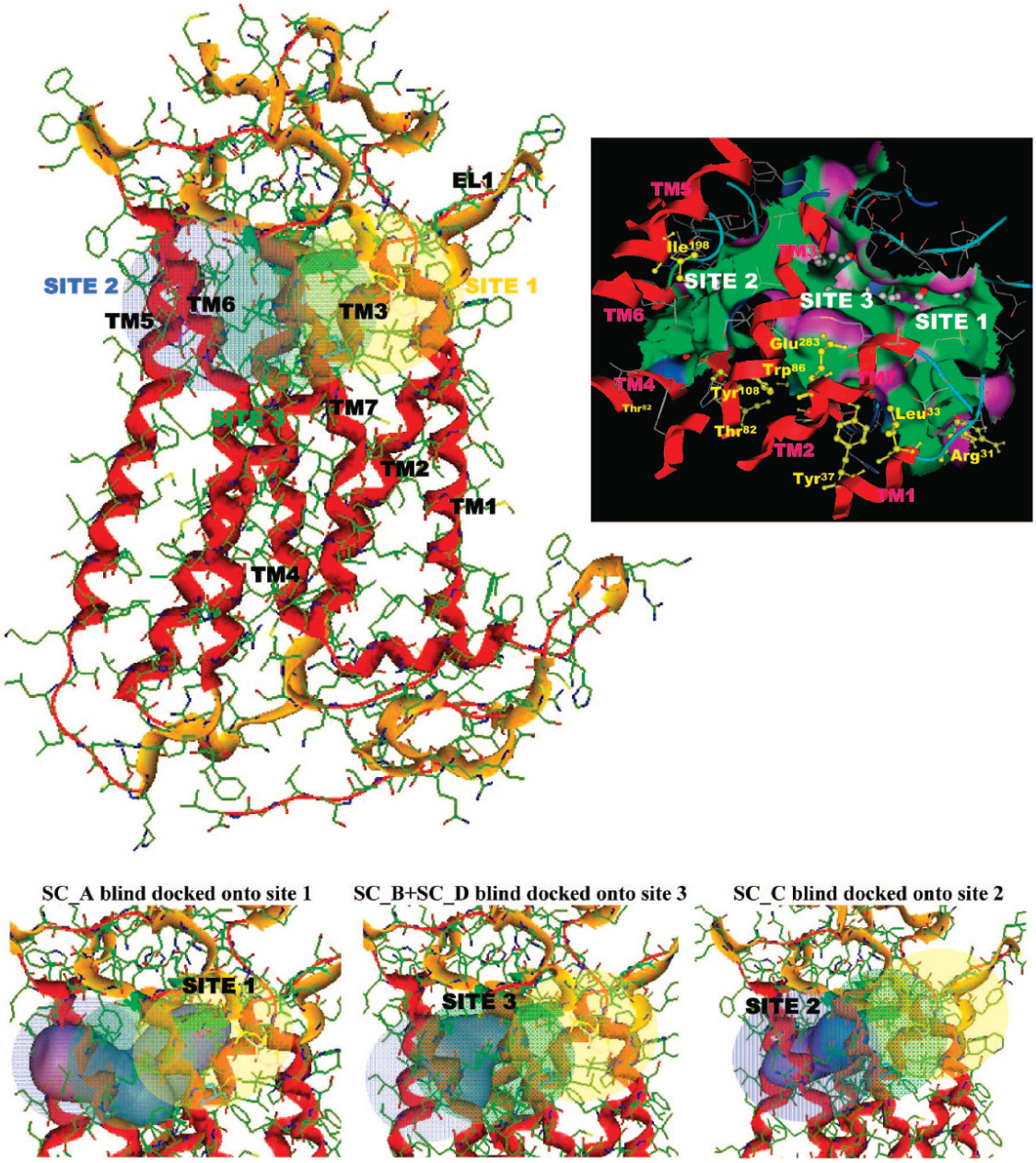


Figura 2.14 Subsitios en la cavidad de unión del CCR5 propuestos por el cribado virtual *consensus shape*. La pseudomolécula *super consensus* A se sitúa en un primer subsitio (Sitio 1) tras el *docking*, delimitado por las transmembranas 1, 2, 3, y 7. La pseudomolécula *super consensus* C se sitúa en un segundo subsitio (Sitio 2) tras el *docking*, delimitado por las transmembranas 3, 5, y 6. Las pseudomoléculas *super consensus* B y D se sitúan en un tercer subsitio (Sitio 3) tras el *docking*, delimitado por las transmembranas 3, 6, y 7, el cual solapa los sitios de unión de las pseudomoléculas A y C. La imagen superior derecha muestra el modelo para el CCR5 con las regiones de unión propuestas marcadas. Arriba a la izquierda, se muestra el mapa de superfície de interacción de van der Waals de la cavidad del receptor CCR5 coloreado según se trate de regiones con enlaces por puente de hidrógeno (lila), hidrofóbicas (verde), y polares (azul). Las transmembranas (TMs) y los residuos importantes para la unión de ligandos que delimitan las tres regiones de unión identificadas se muestran en rojo y amarillo, respectivamente. La fila inferior de imágenes muestra vistas ampliadas de las pseudomoléculas super consenso en los tres subsitios propuestos.

2.3. Cribado Virtual Prospectivo

Una vez validadas las diferentes aproximaciones de cribado virtual mediante el análisis retrospectivo expuesto y establecidos los mejores parámetros para cada una de ellas, se procede al análisis prospectivo de los compuestos de la quimioteca virtual diseñada por S. Pettersson en colaboración con los laboratorios de síntesis y diseño molecular del GEM (Grup d'Enginyeria Molecular del IQS). Con ello, se pretende establecer un *ranking* de posibles nuevos inhibidores de CXCR4. Las moléculas seleccionadas en las primeras posiciones de las listas de *hits* post-procesado por la mayoría de filtros aplicados son las candidatas a ser sintetizadas.

La quimioteca virtual diseñada en nuestro grupo preservando las características principales del AMD3100 consta de 66 compuestos amino/hidrazono-amino/hidrazono, 11 compuestos amino/hidrazono-aldehido y 11 compuestos ciclamo-amino/hidrazono (véase Figura 2.15), algunos de los cuales ya han sido sintetizados por el laboratorio de síntesis del Instituto Químico de Sarriá [34] (véase Artículo VI). Se utiliza pues, para el análisis prospectivo, una base de datos de 34 compuestos no sintetizados todavía pertenecientes a dicha quimioteca virtual cribada junto a compuestos tipo fármaco extraídos de *Maybridge* con propiedades 1D similares a las de los compuestos de la quimioteca (1462 para los filtros farmacofóricos y 4696 para las aproximaciones de *docking* y *shape matching*). Las características de esta base de datos, así como los resultados del cribado que con ella se obtienen se encuentran en el Artículo III.

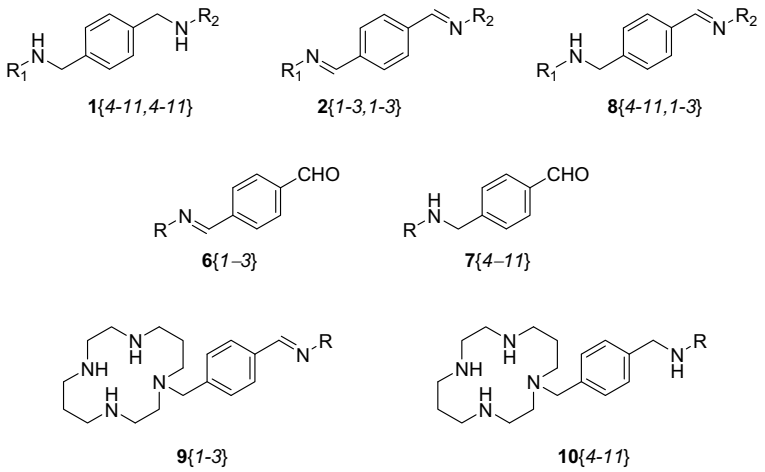


Figura 2.15 Ejemplos representativos de compuestos de la quimioteca combinatoria virtual. Los compuestos **1** son diaminas simétricas (R1 = R2) o no simétricas (R1 ≠ R2); los compuestos **2** son dihidrazonas simétricas (R1 = R2) o no simétricas (R1 ≠ R2); los compuestos **8** son aminohidrazonas; los compuestos **6** y **7** son hidrazonobenzaldehidos y aminobenzaldehidos, respectivamente, y los compuestos **9** y **10** son hidrazono o amino monociclamos sustituidos.

En general, las diferentes aproximaciones utilizadas seleccionan moléculas similares en los primeros porcentajes de las listas de *hits* post-procesado. Los compuestos seleccionados por las diferentes herramientas de cribado virtual *ligand-based* son prácticamente los mismos, mientras que los seleccionados por las herramientas *structure-based docking* incluyen también algunos otros. Las aproximaciones de *docking* Consesus scoring (Autodock Docked Energy, Gold GoldScore y ChemScore), Fred consensus, Hex Docked Energy y Chemgauss3, las aproximaciones *ligand-based* Rocs shape Tanimoto, Parafit shape Tanimoto, Parafit consensus shape Tanimoto, Rocs combo score, Hex shape Tanimoto, y los farmacóforos seleccionan prácticamente las mismas

moléculas en los primeros porcentajes de base de datos cribada. Aunque los filtros *ligand-based* obtienen mejores resultados que las técnicas *structure-based docking* para ambos análisis retrospectivo y prospectivo, los modelos farmacofóricos y Autodock Docked Energy son las herramientas que correlacionan mejor con los datos experimentales. Un *consensus scoring rank by vote* de todos los compuestos encontrados en las primeras posiciones de las listas de *hits* post-procesado ha llevado a seleccionar 5 compuestos para ser sintetizados (véase Figura 2.16), las actividades de los cuales han sido testadas mostrando valores entre 4.1 y 0.022 μg/ml. El Apéndice VIII (póster presentado en Virtual Discovery. Computer-Aided Drug Design and Screening) muestra un resumen gráfico de dichos resultados.

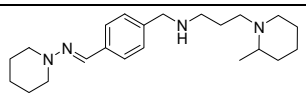
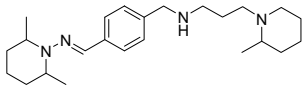
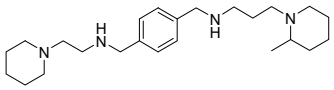
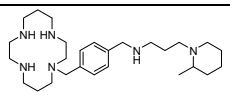
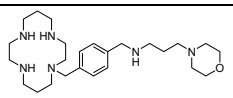
Compound	Name	EC_{50} / μg/ml	CC_{50} / μg/ml
	**8**{*1*,*8*}	> 4.1	4.1
	**8**{*2*,*8*}	0.6	14.6
	**1**{*7*,*8*}	0.4	> 25
	**10**{*8*}	0.022	> 25
	**10**{*11*}	0.058	> 25

Figura 2.16 Resumen de los cinco *hits* seleccionados por el cribado virtual. EC_{50} indica actividad anti-VIH, y CC_{50} es el valor de toxicidad (μg/ml).

2.4. *De Novo Design Structure-Based*

Se ha trabajado en el diseño de nuevos ligandos inhibidores de la unión del VIH al coreceptor CXCR4. Dada la inexistencia de datos cristalográficos para CXCR4, el *de novo design structure-based* se ha llevado a cabo utilizando la estructura modelada expuesta en la Sección 2.1.

Se ha creado una quimioteca de compuestos anti-VIH a partir de los ligandos diseñados *de novo*. Las técnicas de cribado virtual expuestas previamente (véase Sección 2.2) se han utilizado para cribar los compuestos de la quimioteca y seleccionar los candidatos a ser sintetizados y testados.

Dicha quimioteca contiene compuestos diseñados *de novo structure-based* mediante dos procedimientos: búsqueda de fragmentos que encajen en el mapa del sitio activo del receptor y fragmentos que se unan a un *scaffold* dado situado ya en el mapa del receptor. Asimismo, se han derivado compuestos sintéticamente más accesibles en colaboración con el laboratorio de síntesis del Instituto Químico de Sarriá a partir de los resultados obtenidos del *de novo design*. Ambos procedimientos se han llevado a cabo mediante Discovery Studio, accediendo remotamente al CESCA. La metodología utilizada así como los resultados obtenidos se detallan en el Apéndice III. Cabe destacar que las diferentes aproximaciones utilizadas para el cribado virtual han seleccionado moléculas similares en los primeros porcentajes de la lista de *hits* post-procesado. Se han calculado diferentes descriptores y propiedades ADME para estos compuestos utilizando

Qikprop y se han llevado a cabo búsquedas de similitud respecto a un activo conocido mediante Qiksim. Los compuestos con las mejores propiedades ADME o más similares a las de los compuestos activos conocidos han sido seleccionados para ser sintetizados.

2.5. *Fingerprints* de interacción como complemento a estudios de *docking*

Los *fingerprints* de interacción (IFs) han sido desarrollados recientemente con el fin de mejorar la representación y el análisis de las interacciones proteína-ligando. Especialmente han demostrado ser muy útiles en el post-procesado de *outputs* de *docking* como filtro en cribados virtuales y como detección de modos de unión. Dichos métodos se han desarrollado para superar las deficiencias conocidas en lo que respecta a la identificación precisa de conformaciones cercanas a los modos de unión cristalográficos [237, 282]. O. Rabal en su tesis doctoral realizada en el laboratorio de diseño molecular del Instituto Químico de Sarriá ha desarrollado un nuevo *fingerprint* llamado APIF (*Atom Pairs based Interaction Fingerprint*) [35], el cual codifica los modos de unión proteína-ligando en una cadena de *bits* basada en el concepto de pares atómicos. APIF codifica rangos de distancias entre dos puntos de interacción ligando-receptor. En función del tipo de contacto que establece un par de átomos proteína-ligando, estos átomos de la proteína y el ligando se anotan respectivamente como dador de puente de hidrógeno, aceptor de puente de hidrógeno o hidrofóbico. Se analizan todos los posibles pares de interacciones detectados, clasificándose en función de la naturaleza de la interacción (véase Figura 2.17).

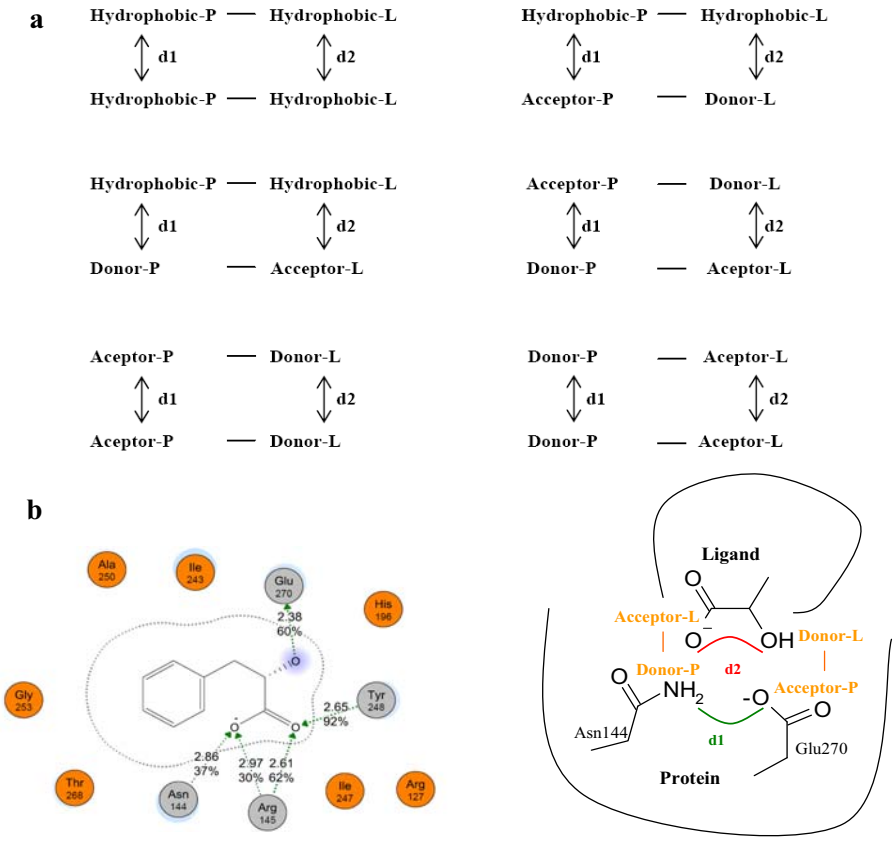


Figura 2.17 *Atom Pairs based Interaction Fingerprint* (APIF). (a) Seis posibles combinaciones de pares de interacciones que definen un conjunto de 49 bits (7 bits para un total de 7 distancias). (b) Codificación del par de interacciones a partir de las distancias medidas entre los dos átomos de la proteína (d1) y los dos átomos del ligando (d2) que interaccionan.

En esta tesis se ha validado el *fingerprint* APIF en un cribado virtual *docking* frente a cinco dianas terapéuticas cristalográficas: tripsina, rinovirus, proteasa VIH, carboxipeptidasa, y el receptor α de estrógeno. Los resultados se comparan por medio de curvas de enriquecimiento con los obtenidos aplicando una implementación del *fingerprint* CHIF introducido por Mpamhanga, que fue validado originalmente en el cribado virtual retrospectivo de antagonistas del receptor α de estrógeno [283]. Se obtienen resultados comparables utilizando cualquiera de los dos *fingerprints* de interacción, mejorando sustancialmente los factores de enriquecimiento con respecto al uso únicamente del *docking* (función de *scoring* GoldScore) en el cribado virtual (véase Figura 2.18). Asimismo se realiza un análisis de modos de unión con el fin de estudiar el mejor método para seleccionar conformaciones con un modo de unión más similar al complejo de referencia cristalizado (véase Figura 2.19). Los resultados muestran que las primeras conformaciones seleccionadas por los *fingerprints* de interacción poseen un modo de unión más similar al complejo de referencia que las conformaciones extraídas del *docking* (GoldScore *scoring function*).

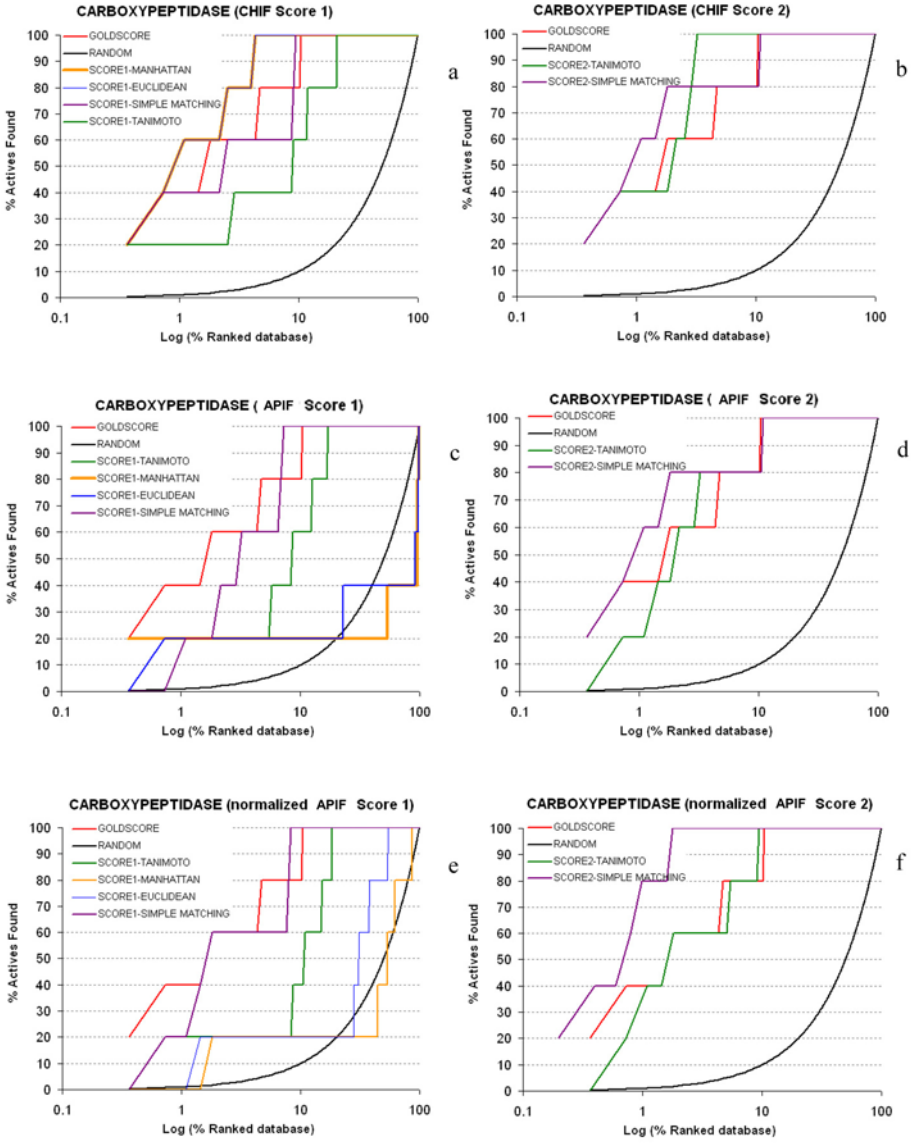


Figura 2.18 Curvas de enriquecimiento obtenidas para carboxipeptidasa (a) implementación CHIF y SCORE1, (b) implementación CHIF y SCORE2, (c) APIF y SCORE1, (d) APIF y SCORE2, (e) APIF normalizado y SCORE1, (f) APIF normalizado y SCORE2. El eje de abcisas es la escala logarítmica del porcentaje de base de datos cribada y el eje de ordenadas corresponde al porcentaje de activos encontrados.

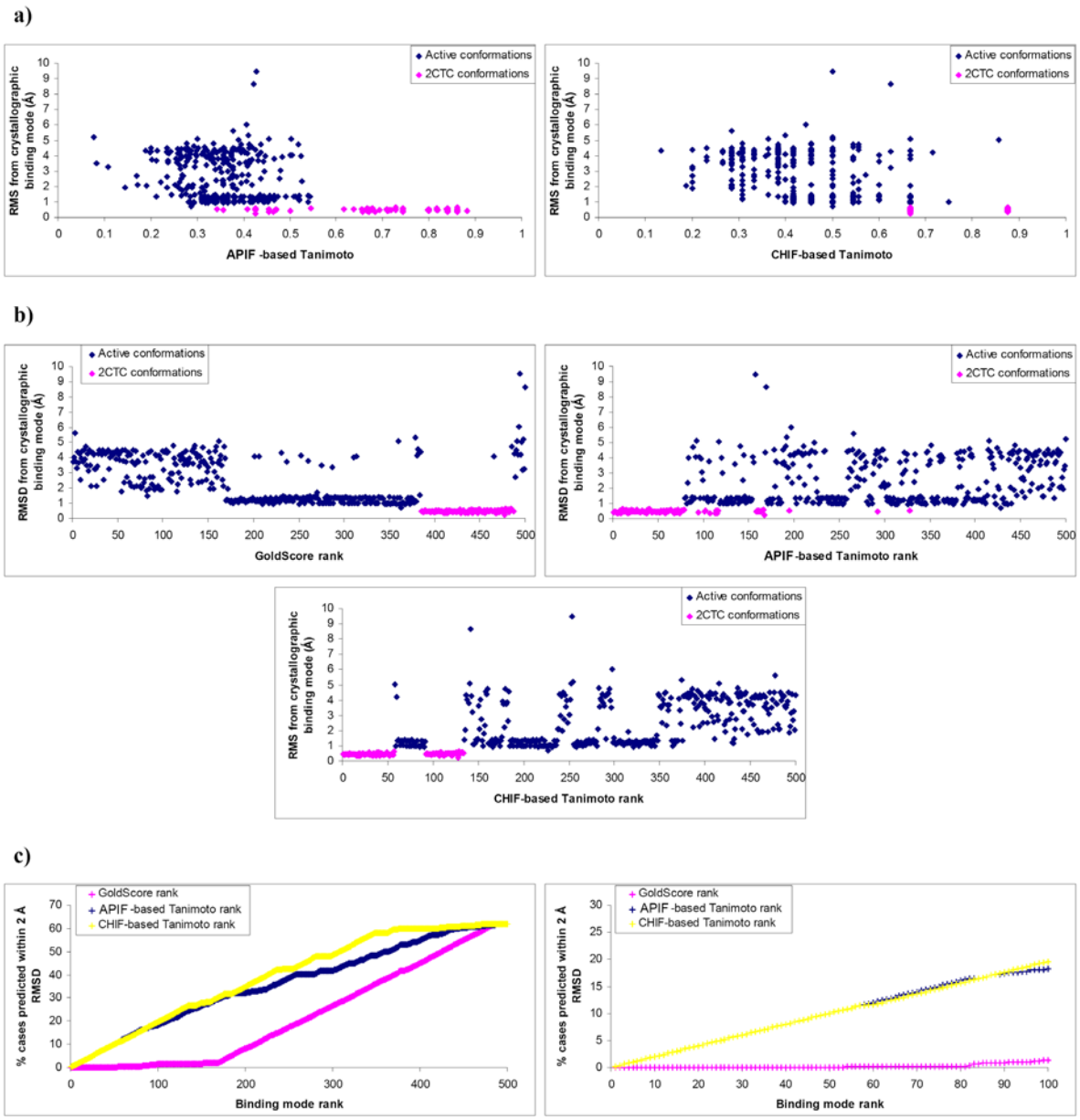


Figura 2.19 Análisis de modo de unión para carboxipeptidasa. a) RMSD respecto al modo de unión cristalográfico (Å) vs APIF-*based* Tanimoto (izquierda) y CHIF-*based* Tanimoto (derecha) para un *set* de 5 ligandos activos. b) RMSD respecto al modo de unión cristalográfico (Å) vs GoldScore *rank* (izquierda), APIF-*based* Tanimoto *rank* (derecha) y CHIF-*based* Tanimoto *rank* (centro). c) % casos con RMSD < a 2Å vs *ranking* del modo de unión para GoldScore (curva rosa), APIF (curva azul), y CHIF (curva amarilla) para las 500 conformaciones activas obtenidas en el *docking* (izquierda) y las 100 primeras ranqueadas (derecha).

En resumen, el estudio demuestra que APIF es útil para ranquear y filtrar resultados de un cribado virtual *docking*. Sin embargo, la calidad de los factores de enriquecimiento obtenidos por APIF depende también del éxito que tenga el *docking*. Los resultados muestran que si el *docking* rinde buenos resultados, entonces APIF proporciona buenos enriquecimientos, mejorando los resultados obtenidos utilizando únicamente una función de *scoring* de *docking*. Por lo tanto, el uso de APIF es una buena manera de seleccionar poses/*hits* virtuales que satisfagan una interacción de referencia proteína-ligando y puede considerarse útil como filtro en el cribado virtual prospectivo de nuevas dianas terapéuticas. El Artículo IV muestra estos resultados así como la metodología utilizada para la validación del *fingerprint*.

3. Artículos

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

Artículo I. Pérez-Nueno, V. I.; Ritchie, D. W.; Rabal, O.; Pascual, R.; Borrell, J. I.; Teixidó, J. Comparison of Ligand-Based and Receptor-Based Virtual Screening of HIV Entry Inhibitors for the CXCR4 and CCR5 Receptors Using 3D Ligand Shape-matching and Ligand-Receptor Docking. *J. Chem. Inf. Model.* **2008**, *48*, 509-533.

Artículo II. Pérez-Nueno, V. I.; Ritchie, D. W.; Borrell, J. I.; Teixidó, J. Clustering and classifying diverse HIV entry inhibitors using a novel consensus shape based virtual screening approach: Further evidence for multiple binding sites within the CCR5 extracellular pocket. *J. Chem. Inf. Model.* **2008** , *48*, 2146-2165.

Artículo III. Pérez-Nueno, V. I.; Pettersson, S.; Ritchie, D. W.; Borrell, J. I.; Teixidó, J. Discovery of Novel HIV Entry Inhibitors for the CXCR4 Receptor by Prospective Virtual Screening. *J. Chem. Inf. Model.* **2009** , *49*, (*Accepted, ci-2008-00468q*).

Artículo IV. Pérez-Nueno, V. I.; Rabal, O.; Borrell, J. I.; Teixidó, J. APIF: A New Interaction Fingerprint Based on Atom Pairs and its Application to Virtual Screening. *J. Chem. Inf. Model.* **2009**, *49*, (*Accepted, ci-2009-00043r*).

Los siguientes artículos, si bien no constan de primera autoría, son el resultado de colaboraciones mantenidas en el período de formación doctoral.

Artículo V. Carrieri, A.; Pérez-Nueno, V. I.; Fano, A.; Pistone, C.; Ritchie, D. W.; Teixidó, J.; Biological profiling of anti-HIV agents and insights into CCR5 antagonist binding using *in silico* techniques. *ChemMedChem* **2009**, *4,* (*Accepted, cmdc.200900010*).

Artículo VI. Pettersson, S.; Pérez-Nueno, V. I.; Ros-Blanco, L.; Puig de la Bellacasa, R.; Rabal,O.; Batllori, X.; Clotet, B.; Clotet-Codina, I.; Armand-Ugón, M.; Esté, J.; Borrell, J. I.; Teixidó, J. Discovery of novel non-cyclam polynitrogenated CXCR4 coreceptor inhibitors. *ChemMedChem* **2008**, *3*, 1549 – 1557.

Artículo I

Comparison of Ligand-Based and Receptor-Based Virtual Screening of HIV Entry Inhibitors for the CXCR4 and CCR5 Receptors Using 3D Ligand Shape Matching and Ligand−Receptor Docking

Violeta I. Pérez-Nueno,[†] David W. Ritchie,*[,‡] Obdulia Rabal,[†] Rosalia Pascual,[†] Jose I. Borrell,[†] and Jordi Teixidó*[,†]

Grup d'Enginyeria Molecular, Institut Químic de Sarriá (IQS), Universitat Ramon Llull, Barcelona, Spain, and Department of Computing Science, King's College, University of Aberdeen, Aberdeen, United Kingdom

Received November 14, 2007

HIV infection is initiated by fusion of the virus with the target cell through binding of the viral gp120 protein with the CD4 cell surface receptor protein and the CXCR4 or CCR5 co-receptors. There is currently considerable interest in developing novel ligands that can modulate the conformations of these co-receptors and, hence, ultimately block virus-cell fusion. This article describes a detailed comparison of the performance of receptor-based and ligand-based virtual screening approaches to find CXCR4 and CCR5 antagonists that could potentially serve as HIV entry inhibitors. Because no crystal structures for these proteins are available, homology models of CXCR4 and CCR5 have been built, using bovine rhodopsin as the template. For ligand-based virtual screening, several shape-based and property-based molecular comparison approaches have been compared, using high-affinity ligands as query molecules. These methods were compared by virtually screening a library assembled by us, consisting of 602 known CXCR4 and CCR5 inhibitors and some 4700 similar presumed inactive molecules. For each receptor, the library was queried using known binders, and the enrichment factors and diversity of the resulting virtual hit lists were analyzed. Overall, ligand-based shape-matching searches yielded higher enrichments than receptor-based docking, especially for CXCR4. The results obtained for CCR5 suggest the possibility that different active scaffolds bind in different ways within the CCR5 pocket.

INTRODUCTION

Acquired Immune Deficiency Syndrome (AIDS) has become a deadly global disease. According to the World Health Organization, some 39 million people now have AIDS and there were 4.3 million new cases in 2006.[1−3] The principal aetiological cause of AIDS is infection of host cells by the human immunodeficiency virus (HIV). Current antiretroviral therapies (ARTs) against AIDS are generally based on reverse transcriptase inhibitors and protease inhibitors. Such therapies can control the spread of the virus and can lead to improved quality of life in patients, but they cannot eliminate the virus from the body and can have undesirable side effects. Several investigators have recognized that one very promising possible alternative approach would be to develop novel therapeutics that can prevent the entry of HIV type 1 (HIV-1) into its target cells and, hence, block the first crucial step of the infection process.[4−6] Following the discovery that HIV infection is initiated by fusion of the virus with the target cell through binding of the viral gp120 protein with the CD4 receptor protein and its co-receptors CCR5 and CXCR4, there has been considerable interest in developing novel ligands that can modulate the co-receptor conformations and, hence, ultimately block virus-cell fusion.[7−11]

Several different computational and experimental approaches are currently being used to identify active compounds against the CXCR4/CCR5 co-receptors.[12−14] Generally, the objective of these approaches is to screen large numbers of candidate drug compounds rapidly. Currently, such computational approaches are often referred to as "virtual screening." Virtual screening has recently become an approximate but useful alternative to laboratory-based high-throughput screening methods for large libraries of compounds. In virtual screening, compounds may be selected and filtered by performing two-dimensional (2D) or three-dimensional (3D) similarity searches, by applying diversity analysis techniques, and by computational docking against the target protein. Compounds may also be selected based on their predicted physical properties (e.g., administration, distribution, metabolism, excretion and toxicity, i.e., ADMET considerations) and their synthetic accessibility. Virtual screening methods may be classified as "structure-based methods" (e.g., docking), which are used when the structure of the receptor is known or can be modeled, and "ligand-based methods", in which the screening process is based only on the characterization of known active compounds (e.g., by constructing pharmacophoric models or by performing quantitative structure−activitiy relationship (QSAR) studies). For example, Afantitis et al.[12] and Aher et al.[13] recently performed QSAR and virtual screening studies of CCR5 antagonists derived from 1-(3,3-diphenylpropyl)-piperidinyl amides. These studies show that the key chemical and

* Author to whom correspondence should be addressed. Tel.: +44 1224 272282. Fax: +44 1224 273422. E-mail: dritchie@csd.abdn.ac.uk.

[†] Grup d'Enginyeria Molecular, Institut Químic de Sarriá (IQS), Universitat Ramon Llull.

[‡] Department of Computing Science, King's College, University of Aberdeen.

10.1021/ci700415g CCC: $40.75
Published on Web 02/26/2008

structural requirements for high-affinity binders can be identified using ligand information such as one-dimensional (1D) physicochemical properties, two-dimensional (2D) topological descriptors, and three-dimensional (3D) properties such as steric, electrostatic, hydrophobic, and hydrogen bond acceptor/donor fields around a family of aligned molecules.

In principle, structure-based methods might be expected to give better results than ligand-based approaches, because they try to model the physics of protein−ligand interactions. For example, Kellenberger et al.[14] used a combination of 2D and 3D structure-based screening techniques to identify 10 CCR5 binders from a library of 1.6 million compounds. However, in the final high-throughput docking stage, they found that the two different docking algorithms that were used produced very few common hits, and that only a handful of these shared compatible poses. Thus, the results of structure-based approaches can be seen to depend critically on the quality of the protein structure and docking protocol applied. In general, both ligand-based and structure-based approaches inevitably have fundamental limitations. For example, simply finding the best way to superpose related ligands remains an open problem. Furthermore, in high-throughput structure-based approaches, it is generally impractical to include an explicit solvation model or to fully cover the conformational spaces of each receptor−ligand pair, because of the high computational cost of performing molecular dynamics (MD) simulations for each putative complex. Indeed, if the target protein must be model-built, its initial 3D structure will very likely contain structural errors. Nonetheless, despite such limitations, from a purely utilitarian point of view, it is still possible to use, e.g., enrichment plots to compare objectively the relative abilities of ligand-based and structure-based approaches to identify known binders through retrospective virtual screening studies, and to use the knowledge gained to predict rationally new potential actives.

This article describes, for the first time, a thorough comparison of the utility of ligand-based and docking-based virtual screening approaches to find entry blockers for the CCR5 and CXCR4 co-receptors, taking into account multiple known families of active compounds for each target. To achieve this, a database consisting of 602 compounds which are known to be active inhibitors of CXCR4 and CCR5 was assembled from the literature. This database consists of 13 families of CCR5 inhibitors and 5 families of CXCR4 inhibitors, including well-known ones such as TAK-779, SCH-C and their derivatives for CCR5, and the bicyclam series for CXCR4.[11] A similar database of some 4700 presumed inactive compounds with 1D properties that are comparable to those of the actives was also compiled to provide decoys for the virtual screening protocols. These datasets constitute a valuable resource for virtual screening studies of the CXCR4 and CCR5 co-receptors. Because no crystal structures of CCR5 and CXCR4 currently exist, 3D models were built by homology from the nearest available template structure, bovine rhodopsin.[15,16] A variety of docking programs were used to perform structure-based screening of the database to determine the best methodology for identifying active CXCR4 and CCR5 antagonists. Several ligand-based shape-matching algorithms were also compared, primarily using AMD3100 and TAK779 as query compounds for the CXCR4 and CCR5 co-receptors, respectively, because considerable experimental data and theoretical predictions are available for these ligands.[17−21] Our results show that CXCR4/CCR5 receptor-based and ligand-based approaches can be used to discriminate actives from inactives in a retrospective virtual analysis. We believe that these approaches can now be used in prospective virtual analysis to select candidates for the rational design of HIV-1 entry blockers.

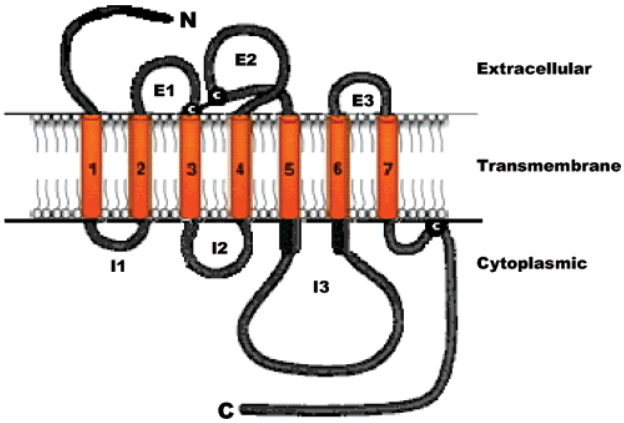


Figure 1. Topology of rhodopsin-like G protein coupled receptors. The three extracellular loops are labeled E1, E2, and E3, and the three intracellular loops are labeled I1, I2, and I3.

METHODS

Construction of CXCR4 and CCR5 Models for Docking. CXCR4 and CCR5 models were built using bovine rhodopsin (PDB code 1HZX) as a template.[15] This 2.8 Å resolution X-ray structure consists of a transmembrane (TM) domain of seven α helices connected by three extracellular loops (ELs) and three intracellular loops (ILs), as illustrated schematically in Figure 1. Bovine rhodopsin is the first solved structure of the G protein-coupled receptor (GPCR) family, and it has been used as a template for modeling many other GPCR drug targets. Generally, GPCR homology modeling starts with the assumption that all GPCR family members share a common topology to bovine rhodopsin, even when the sequence identity is as low as 20%,[16] as is the case for CXCR4 or CCR5. Thus, as a first step, the TM segments for both proteins were predicted using HMMTOP,[22] TMHMM,[23] MEMSAT,[24] and DAS.[24] The results were compared to the 3D model proposed by Gerlach et al.[25] for CXCR4 (SwissProt accession number P30991) and the model constructed by Paterlini et al.[26] for CCR5 (SwissProt accession number P51681). We selected the same definition of TM segments as Gerlach et al. for CXCR4 (i.e., TM1, 39−64; TM2, 75−98; TM3, 109−135; TM4, 152−174; TM5, 196−223; TM6, 236−265; TM7, 279−308) and one very similar to that of Paterlini et al. for CCR5 (i.e., TM1, 34−56; TM2, 69−89; TM3, 101−130; TM4, 142−163; TM5, 191−216; TM6, 230−259; TM7, 277−312), which differs only by one residue in the definition of TM2. The TM segments were aligned with those of the template using the BLOSUM62 and GONNET matrices in MODELLER 6[27] without allowing gaps in the helices and by requiring all highly conserved family A GPCR residues to be aligned. The loop sequences were aligned separately in MOE,[28] using the DAYHOFF matrix with a gap-open penalty of 3.0 and a gap-extension penalty of 1.0. The sequence identity is 21% between

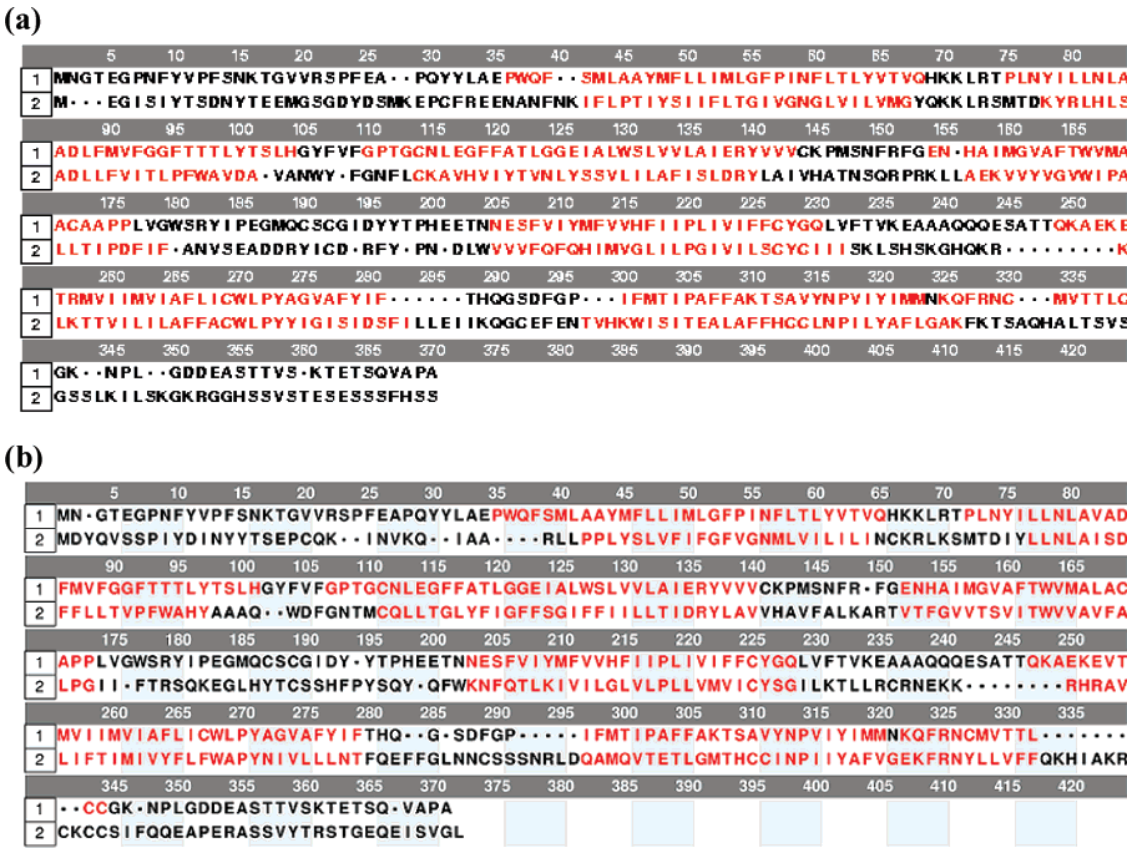


Figure 2. Pairwise sequence alignments of bovine rhodopsin and the CXCR4 and CCR5 co-receptors. Panel a shows bovine rhodopsin (1) and human CXCR4 (2), yielding 21% sequence identity; panel b shows bovine rhodopsin (1) and human CCR5 (2), yielding 20% sequence identity.

CXCR4 and bovine rhodopsin, and 20% in the case of CCR5. It is worth mentioning that the alignments matched the conserved cysteines Cys110 and Cys187 of bovine rhodopsin with Cys109 and Cys186 of CXCR4 and with Cys101 and Cys178 of CCR5. This highly conserved disulfide bond appears in all GPCRs and is structurally important in forming the EL2 conformations.[29] Figure 2 shows the final alignments between bovine rhodopsin and CXCR4 and CCR5.

The 3D structures of CXCR4 and CCR5 were built using MODELLER. The disulfide bonds Cys109-Cys186 of CXCR4 and Cys101-Cys178 of CCR5 were modeled from the coordinates of the equivalent disulfide bond in bovine rhodopsin. However, MODELLER did not automatically identify the second disulfide bond of both CXCR4 (Cys28-Cys274) and CCR5 (Cys20-Cys269); therefore, these regions were modeled ab initio. Forty different models with root-mean-square (rms) deviations of up to ±4 Å were generated from different randomized initial coordinates and were optimized. The model with the lowest target function was then selected and refined by simulated annealing. For brevity, these CXCR4 and CCR5 structures will subsequently be called "NO-loops" models.

Loops were modeled first by homology modeling with bovine rhodopsin, starting from the DAYHOFF alignment. The intermediate loop regions corresponding to segments I1 (65−74), E1 (99−108), I2 (136−151), I3 (224−235), E3 (264−281) for CXCR4, and I1 (217−229), E1 (260−276), I2 (131−141), I3 (56−68), E3 (90−100) for CCR5 were then modeled ab initio with MODELLER.[30] EL2 was initially modeled entirely by homology, because it is a generally conserved motif in GPCR family A.[29] One hundred structures were generated and the one with the lowest objective function was selected and refined by simulated annealing. For brevity, the structures built using this loop-modeling approach will be called "MODELLER-loops" models. Visual inspection and preliminary docking calculations of these models showed that EL2 was not sufficiently open to allow ligand binding. Hence, all loops including EL2 (i.e., residues 175−195 and 165−195 for CXCR4 and CCR5, respectively) were remodelled ab initio by conformational search with CONGEN,[31] using a direct conformational search with disulfide bond constraints and real space renormalization.[32−36] To open the loops, the conformation with the highest rms deviation from the starting one was selected in each case. The loops were then energy-minimized using steepest descent, conjugate gradient, and truncated Newton optimization consecutively, with a maximum number of iterations of 100, 100, and 200, respectively, and a convergence gradient of 1000, 100, and 0.01, respectively. The CXCR4 and CCR5 structures built using this loop-modeling approach will subsequently be called "CONGEN-loops" models. The structural quality of the final models was checked using PROCHECK.[37] This shows that 93.5% of CXCR4 residues and 97.5% of CCR5 residues fall within favored or allowed regions of the Ramachandran map (CXCR4: 64% favored, 29.3% allowed; CCR5: 71.2% favored, 26.3% allowed). Side chain χ_1 and χ_2 torsion angles were found in their most favorable regions without stereochemical conflicts. These statistics are consistent with other homology models of these receptors.[26,38]

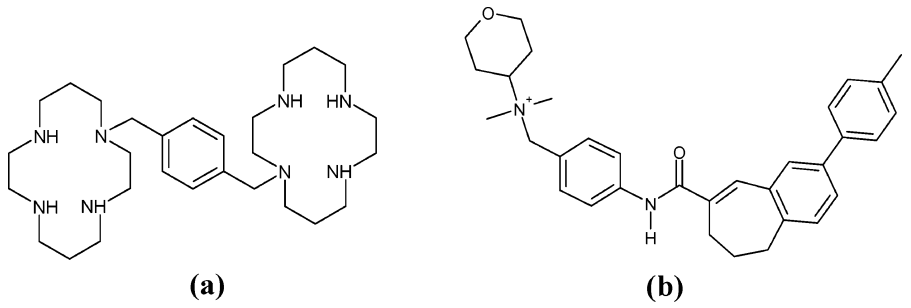


Figure 3. Chemical structures of (a) AMD3100 and (b) TAK779.

Regarding charge assignments for polar residues, a working pH of 7 was assumed. At this pH, and without experimental evidence of protonation of acidic residues or deprotonation of basic ones, all Asp, Glu, Arg, and Lys residues were considered charged. His residues, having a pK_a of 6.6, were considered neutral. Polar hydrogens were added to each protein, and their positions were relaxed using the AMBER force field in MOE with Amber99[39] partial charges.

Receptor Binding-Site and Binding-Mode Analyses. It is reasonable to assume that if a good 3D model of a receptor can be built, then an automatic docking protocol should be able to locate a high affinity ligand within the binding site correctly. Hence, to validate our receptor models, blind docking was initially performed with AUTODOCK.[40] Ligand binding within the site-directed mutagenesis (SDM) defined binding pocket was subsequently analyzed in detail using AUTODOCK and GOLD.[41] AMD3100 and TAK779 inhibitor structures were built, assigned Gasteiger partial charges,[42] and minimized in MOE with the MMFF94 force field. AMD3100 was docked against CXCR4, and TAK779 against CCR5, respectively. The structures of these ligands are shown in Figure 3. The results obtained were assessed using knowledge of the SDM data. For example, mutagenic substitutions of 16 CXCR4 amino acids located in TM helices TM3, TM4, TM5, TM6, and TM7 have identified three acidic residues—Asp171, Asp262, and Glu288—as the main interaction points for AMD3100 binding.[25,43−49] Two of these residues (Asp262 of TM6 and Glu288 of TM7) are in one extreme, whereas the third (Asp171 in TM4) is in the opposite extreme of the ligand binding pocket. Similarly, mutagenic substitutions in CCR5 have implicated Glu283, Trp86, Tyr37, Tyr108, Leu33, Val83, Ala90, and Gly286 (located within TM helices TM1, TM2, TM3, and TM7) as comprising the binding site for the TAK779, AD101, and SCH−C ligands. These results also suggest that Glu283 acts as a counterion for the positively charged N atom common to the TAK779, AD101, and the SCH−C ligands.[50−53]

For the AUTODOCK blind docking experiment, a 181 × 181 × 181 grid with a grid spacing of 0.375 Å was used, centered on the SDM-defined ligand binding site. This grid enclosed the entire protein structure, with the ligand initially placed far from the protein, to avoid excluding the possibility of finding other binding sites. A smaller (61 × 61 × 61) grid was used for the subsequent binding mode analysis calculations. In each case, 100 independent Lamarckian genetic algorithm (LGA) runs were performed and pseudo-Solis and Wets minimization methods were applied using default parameters. Each docking run was repeated five times.

Table 1. Families of CXCR4 and CCR5 Inhibitors Compiled in the Current Study

family	number of compounds	references
CXCR4 Inhibitors		
tetrahydroquinolinamines	123	11, 54−58
KRH derivatives	23	11, 59−62
macrocycles	4	63
AMD derivatives	94	11, 63−68
cyclic peptides	2	69
other	2	70
total	**248**	
CCR5 Inhibitors		
SCH derivatives	120	71−73
diketopiperazines	9	74−78
anilide piperidine N-oxides	22	79
AMD derivatives	3	68
4-piperidines	10	80, 81
4-aminopiperidine or tropanes	26	80, 82, 83
1,3,4-trisubstituted pyrrolidinepiperidines	9	84
phenylcyclohexilamines	9	85−90
TAK derivatives	66	91, 92
1-phenyl-1,3-propanodiamines	57	93−95
1,3,5-trisubstituted pentacyclics	10	96
N,*N*′-diphenylureas	4	97
5-oxopyrrolidine-3-carboxamides	5	98
other	4	99
total	**354**	

For the GOLD binding mode docking runs, the ligand binding site was limited to all protein atoms within 20 Å from the centroid of the SDM-defined binding residues. The GOLD cavity-detection algorithm was used to confine the calculation to concave regions in the vicinity of the binding site. GOLD uses a genetic algorithm (GA) to explore the possible binding modes. As with AUTODOCK, 100 docking runs per experiment were performed, with each run consisting of a maximum of 100 000 GA operations. All other GA parameters used default values. Cutoff distances of 2.5 Å for hydrogen bonds and 4.0 Å for nonbonded contacts were set.

Virtual Screening Data Preparation. We compiled a large set of 248 CXCR4 and 354 CCR5 antagonist inhibitors from the literature, which mainly consists of 5 representative families of CXCR4 inhibitors and 13 representative families of CCR5 inhibitors, as listed in Table 1. In the compiled dataset, 94% of the molecules have activity values of <0.1 μM against CXCR4 or CCR5. The remaining 6% have activities in the range of 0.1−1 μM (4%), 1−10 μM (1%), and 10−100 μM (1%). Figure 4 shows some representative members of each family.

To avoid potential bias of the virtual screening results due to large differences in basic properties (molecular weight,

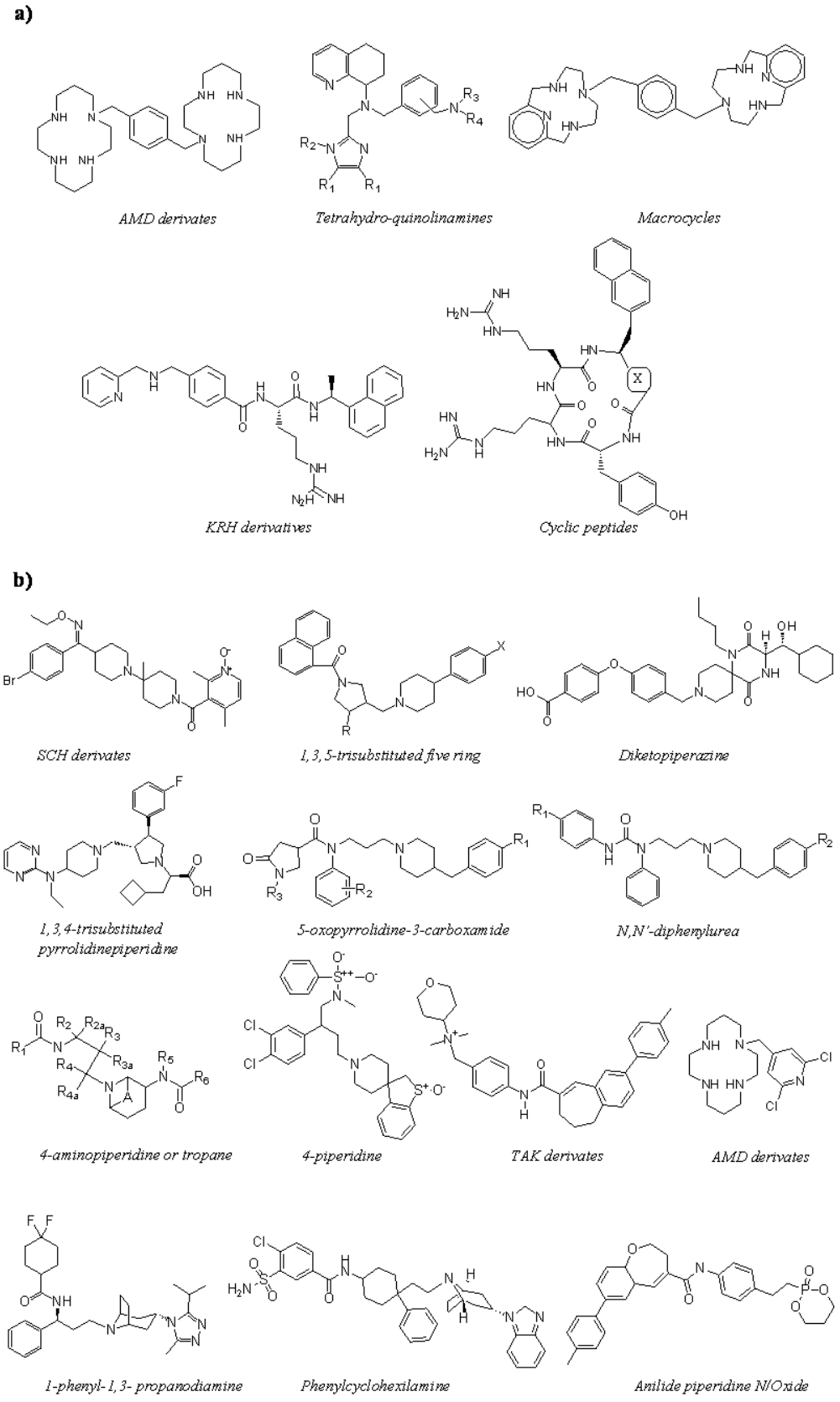


Figure 4. Representative structures of (a) 5 families of CXCR4 inhibitors, and (b) 13 families of CCR5 inhibitors.

Table 2. Summary of the 1D Physicochemical Properties of Active and Inactive Molecules in the Screening Database[a]

molecule	molecular weight	number of rotatable single bonds, b_1rotN	number of hydrogen-bond acceptor atoms, a_acc	number of hydrogen-bond donor atoms, a_don	number of hydrophopic atoms, ahyd	octanol−water partition coefficient, SlogP
248 CXCR4 actives	507.3 (74.4)	9.2 (4.9)	4.9 (1.1)	1.7 (1.3)	27.6 (4.2)	4.3 (3.0)
354 CCR5 actives	559.6 (86.8)	12.9 (4.1)	4.7 (1.6)	0.9 (0.8)	28.9 (4.2)	5.7 (1.4)
4696 inactives	497.4 (45.6)	6.2 (2.4)	3.6 (1.6)	0.9 (1.0)	21.8 (4.1)	5.5 (1.9)

[a] The average value is given in normal font, and the standard deviation is given in parentheses.

etc.), it is important to work with focused screening libraries.[100] Hence, a set of 4696 presumed inactive compounds was assembled from the Maybridge Screening Collection database[101] in such a way that several 1D properties were similar to those of the active compounds (molecular weight, number of rotatable single bonds, number of hydrogen-bond acceptor atoms, number of hydrogen-bond donor atoms, octanol−water partition coefficient, and number of hydrophobic atoms). Table 2 shows that the average and standard deviations of these properties calculated by MOE[28] are quite similar for the active and inactive pools. The 3D structures of all ligands also were built using MOE. The structures were protonated at physiological pH (i.e., pH 7), Gasteiger partial charges were assigned, and the geometry was optimized using the MMFF94 force field. The ligands were then approximately located into their respective receptor binding pockets with the MOE FlexAlign module,[102] using the docked TAK779 and AMD3100 conformations as superposition templates for CCR5 and CXCR4, respectively.

Docking-Based Virtual Screening. Docking-based screening against CXCR4 and CCR5 was performed using AUTODOCK 3.0,[40] GOLD 3.0.1,[41] FRED 2.2.1,[103] and HEX 4.8.[104,105] In AUTODOCK, 10 independent LGA runs were performed, using the same protocol as that applied in the binding mode analysis described previously. The Docked Energy scoring function was used to rank the ligand databases. In GOLD, 10 independent GA runs were performed, also using the same protocol as that applied in the binding mode analysis calculations, but allowing early termination when the top three docking solutions for each ligand were within 1.5 Å of each other. Protein hydrogen bond constraints with a weighting factor of 10 were specified to obtain binding modes that involved key SDM-defined binding residues. In the case of CXCR4, the ligands were constrained to form a hydrogen bond to one of the Glu288, Asp171, or Asp262 carbonyl oxygens. The GoldScore and ChemScore scoring functions were then used to rank the ligand databases. A consensus "Rank-by-Rank"[106] score was also calculated by determining the final rank of every compound in the database as the average rank of the AUTODOCK and the two GOLD scoring functions. In FRED, exhaustive rigid body optimization, pose ranking, and force field refinement were performed using default parameters. The search space was specified using a shape-based site detection algorithm and the position of a given bound ligand (docked AMD3100 and TAK779 for CXCR4 and CCR5, respectively). Because all the ligands in our database had already been flexibly aligned by MOE to the docked AMD3100 and TAK779 conformations, their conformations were suitable to be rigidly docked with FRED. Multiple FRED scoring functions were calculated (PLP, Chemgauss3, Shapegauss, OEChemScore, ScreenScore and ChemScore). A consensus scoring hit list was also calculated over all the selected scoring functions. In HEX, docking was performed using a six-dimensional shape-only superposition correlation search with a translational distance range of 10 Å from the SDM-defined active site center.

Shape-Matching-Based Virtual Screening. Ligand-based virtual screening was performed using PARASHIFT 06,[104,107] ROCS 2.2,[108] and HEX 4.8[104,105] by superposing each of the database compounds onto a given query molecule (i.e., the docked TAK779 and AMD3100 conformations for CCR5 and CXCR4, respectively). PARASHIFT uses two software modules: PARASURF and PARAFIT. PARASURF[107] was used to calculate, from semiempirical quantum mechanics theory, the molecular shape and electronic properties of all ligands and to encode these properties as spherical harmonic (SH) expansions. PARAFIT[104] was then used to superpose every database compound onto the query by exploiting the special rotational properties of the SH expansions. These superpositions used the SH shape Tanimoto as the objective function, which was also used to rank the ligand database. Similarly, HEX 3D shape Tanimoto scores were calculated by maximizing the 3D density overlap between pairs of colocated molecules using default HEX search parameters.

PARASHIFT and HEX superpositions were performed using the conformation computed by the MOE FlexAlign option for each database compound. However, ROCS shape-matching calculations were performed using different conformers to study the influence of different query and database compound conformations. OMEGA[109] was used to calculate 10 further conformations of each query molecule, starting from the docked conformations, as well as 10 different conformations of every compound in the ligand database. Superposition of atom-centered Gaussian functions[108] was then performed with ROCS to compute shape-based overlays of all conformers of every compound in the database using AMD3100 and TAK779 as query molecules in one or more conformations. Database molecules were then ranked according to their shape Tanimoto scores for each query molecule. ROCS was also used in "color optimization" mode to maximize both the shape and chemical property overlays obtained by aligning fragments with similar chemical properties (e.g., proton donor/acceptor, cationic/anionic, and hydrophobicity/aromaticity).

Enrichment Factors. Following the docking and shape-matching calculations, all compounds were sorted into ranked lists based upon their docking and shape-matching scores. These lists were then used to plot the percentage of known actives found versus the percentage of the ranked database screened and to calculate the enrichment factor (EF) at 1%, 5%, and 10% of the screened database. The EF measures the number of known ligands in the top-ranked list, relative to a random selection. Thus, for a library built with the $N_{sampled}$ top compounds of the ranked library, the EF is defined as

$$EF = \frac{(Hits_{sampled}/N_{sampled})}{(Hits_{total}/N_{total})} \quad (1)$$

where $Hits_{sampled}$ correspond to the number of active compounds in the subset ($N_{sampled}$), and $Hits_{total}$ represents the total number of actives in the entire database (N_{total}). It can be observed that the EF has a fixed maximum at any given percentage of the database screened, given by $N_{total}/Hits_{total}$. At 1%, the maximum is 100, at 2% the maximum is 50, and at 10% screened the maximum obtainable enrichment is 10. An EF of 1 corresponds to a random distribution of active molecules in the ranked database.

Some authors have previously noted the importance of the ability to identify diverse chemotypes as a measure of algorithm robustness.[110] In other words, a virtual screening procedure that can retrieve a high number of representative compounds of each scaffold within the first percentages of the database is more desirable than a procedure that gives a high EF obtained by identifying multiple compounds from the same chemical series. Hence, the ability of docking and shape-matching techniques to retrieve a diverse scaffold pool that might facilitate lead structure identification was also assessed. For this purpose, plots were made to determine how many compounds must be screened before at least one member of each active scaffold class is identified.

RESULTS

CXCR4 Binding Site and Binding Mode Analysis. In the blind docking of AMD3100 against CXCR4, AUTODOCK was able to recognize the SDM-defined binding site around Asp171, Asp262, and Glu288 for all repetitions using the NO-loops model, none for the MODELLER-loops model, and three out of five repetitions for the CONGEN-loops model. These results are shown in Figure 5. Visual inspection of this figure suggests that the CXCR4 NO-loops model allows all ligand conformations clash-free access to the binding site, whereas the MODELLER-loops model hinders ligand binding, because of the closed conformation of EL2. However, the more-open EL2 conformation in the CONGEN-loops model is able to accommodate ligand access to the cavity. To quantify these differences systematically, for each docking run, the closest distances from the carboxylic oxygens of the three key binding residues Asp171, Asp262, and Glu288 and any of the eight AMD3100 nitrogens were recorded. Table 3 shows these distances, along with the corresponding AUTODOCK binding energy and the experimentally determined ΔG value.

Considering the AMD3100 binding mode calculations, both AUTODOCK and GOLD were able to find satisfactory binding modes within the SDM-defined cavity. Visual inspection of the resulting poses and consideration of the key interatomic distances listed in Table 3 show that the expected ligand−receptor binding interactions are indeed present in the docked complexes. Hence, our results generally agree with previous computational studies. For example, Schwartz et al.[49] and Sadler et al.[111] (using AMD3100-Zn_2) found a "sandwich" conformation in which two nitrogens of one cyclam ring interact with the two carboxyl oxygens of Asp171 (TM4), and where one face of the other cyclam ring interacts with the two carboxyl oxygens of Asp262 (TM6) and the opposite face interacts similarly with Glu288 (TM7). In the present study, AMD3100 is not complexed with a metal, so our results are more comparable to those of Schwartz et al.[25] and Trent et al.[112] Schwartz et al.[25] found a docked conformation in which two nitrogens of one cyclam ring interact with the two carboxylic oxygens of Asp171 and two nitrogens of the other cyclam ring interact with the two carboxylic oxygens of Asp262. Trent et al.[112] found a docked conformation in which two nitrogens of one cyclam ring interact with the two carboxylic oxygens of Asp262, and two nitrogens of the other cyclam ring interact with the two carboxylic oxygens of Glu288. As before, to analyze the calculated binding modes in more detail, distances were measured between the Asp171, Asp262, and Glu288 carboxylic oxygens and all of the eight AMD3100 nitrogens for the nearest and lowest energy conformations found by AUTODOCK for each of the NO-loops, MODELLER-loops, and CONGEN-loops models. In the latter case, distances were also measured for the conformations obtained using GoldScore. The distances in Tables 4−6 indicate that our binding mode is most similar to that of Trent et al.,[112] because the distances between two nitrogens of one cyclam ring with the two carboxylic oxygens of Asp262 and two nitrogens of the other cyclam ring with the two carboxylic oxygens of Glu288 are shorter than the distances between two nitrogens of one cyclam ring with the two carboxylic oxygens of Asp171. Figure 6 shows the corresponding binding conformation. Furthermore, the CONGEN-loops model gives better results than MODELLER-loops, because EL2 does not cause steric clashes in this case. This suggests that the aforementioned blind docking and binding site docking analyses have validated our CXCR4 model, and that, therefore, its use would be suitable in structure-based virtual screening.

CCR5 Binding Site and Binding Mode Analysis. As with CXCR4, blind docking was performed with the CCR5 NO-loops, MODELLER-loops, and CONGEN-loops models, this time using TAK779 as the high-affinity probe ligand. Previous SDM results show that the substitution of Glu283, Trp86, Tyr37, Tyr108, Leu33, Arg31, Ile198, and Thr82 residues by alanine causes the greatest inhibition of antiviral activity of TAK779.[26] In this case, AUTODOCK was also able to recognize the supposed binding site within the cavity formed by these key SDM residues for all the repetitions with the NO-loops model, none for the MODELLER-loops model, and three out of five repetitions with the CONGEN-loops model. Figure 7 summarizes these results. As with CXCR4, it can be observed that the CCR5 NO-loops and CONGEN-loops structures allow several conformations inside the binding site, whereas the closed conformation of EL2 in the MODELLER-loops model sterically prohibits ligand entry to the binding site. Table 7 lists the key polar and hydrophobic interatomic distances between TAK779 and the key SDM residues, calculated using HBPLUS,[113] along with the AUTODOCK docking energy and experimental ΔG value for this complex.

Analysis of the TAK779 docking binding mode results shows that AUTODOCK was able to recognize the main binding interactions, according to the SDM data. As in the blind docking case, distances were measured between key SDM residues and TAK779 (the nearest conformations and lowest energy conformations found by AUTODOCK) for the three CCR5 receptor models. These results are shown in Tables 8 and 9. Overall, the low distances found between

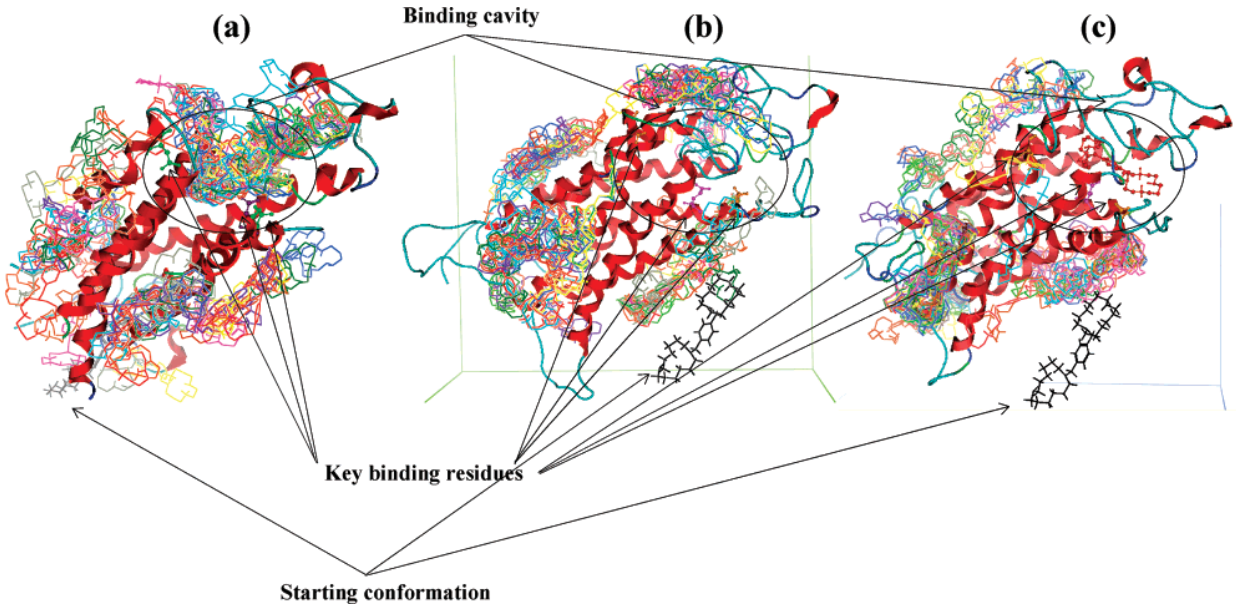


Figure 5. Blind docking results for AMD3100-CXCR4 obtained using AUTODOCK. The starting conformation is shown as black sticks, and the resulting best 100 docked conformations are shown as colored sticks. Key receptor binding site residues are shown as balls and sticks: (a) NO-loops model (many conformations are found within the binding pocket); (b) MODELLER-loops model (closed-loop conformation) (no ligand conformations are found in the binding pocket due to the steric clashes with EL2); and (c) CONGEN-loops model (open-loop conformation; the lowest energy docked conformation, shown as a red ball and stick structure, is found within the binding pocket).

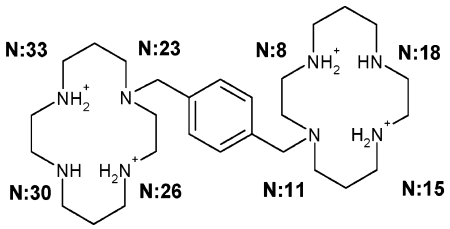
Glu238 and the TAK779 ammonium nitrogen agree with the SDM data,[51,72,114,115] although short distances are also found between Glu283 and TAK779 peptide nitrogen for the lowest energy conformations. In addition, the hydrophobic interactions of TAK779 with Trp86, Tyr37, Tyr108, and Leu33 are more favored than those of TAK779 with Arg31, Ile198, and Thr82 residues. This agrees well with the SDM data, which shows that alanine substitutions of Trp86, Tyr37, Tyr108, and Leu33 strongly inhibit the antiviral activity of TAK779, whereas mutations of Arg31, Ile198, and Thr82 only moderately inhibit TAK779 antiviral activity.[26] As in the CXCR4 case, the CCR5 CONGEN-loops model gives better results than the MODELLER-loops model. Therefore, our results seem to agree with the earlier computational studies of Dragic et al.,[52] Seibert et al.,[50] Zhou et al.,[51] Paterlini et al.,[26] and Fano et al.[116] All of these studies indicate (a) the importance of the interaction between Glu283 and the TAK779 ammonium nitrogen due to the absence of other nearby positively charged counterions; (b) the main hydrophobic interactions existing with Tyr37, Tyr108, Leu33, and Ile198; and (c) the influence of EL2, with respect to ligand binding. Specifically, our binding mode conformation that is shown in Figure 8 is oriented in a manner similar to that of Paterlini et al.,[26] with the TAK779 benzyl pyran ammonium group interacting with helices TM1, TM2, and TM7, with a near contact between ammonium nitrogen of TAK779 and Glu283, and with a hydrophobic interaction with Tyr37. Like the model of Paterlini et al., the methylphenylbenzocycloheptenyl moiety is also buried in the TM barrel and has hydrophobic interactions with Tyr108 and Ile198. These observations support the validity of our CCR5 model and therefore suggest its suitability for its use in structure-based virtual screening.

It is worth mentioning that, using MD, it is possible to refine the CXCR4 and CCR5 docking poses to obtain ligand conformations more similar to the key SDM residues.[26,38,116−119] For example, applying 200 ps of AMBER MD to our docking poses, using the protocol described in Orozco et al.,[120] gives ligand conformations with an average distance of 2 Å closer to the key binding residues. However, performing MD on every ligand in a database as large as ours would currently require a supercomputer.

Docking Enrichments. To analyze the ability of our receptor model structures to discriminate active compounds from decoys, enrichment curves were calculated using AUTODOCK docking energies and FRED PLP, Chemgauss3, Shapegauss, OEChemScore, ScreenScore, ChemScore, and Consensus scoring functions for CXCR4 and CCR5. In addition, enrichment curves were also calculated using the GOLD GoldScore, ChemScore, and Rank-by-Rank Consensus scoring functions for CXCR4. Figure 9 shows the enrichment curves obtained using these docking-based scoring functions. Inspection of these results shows that the enrichments obtained with the AUTODOCK Docked Energy scoring function with CCR5 surpass those with CXCR4. Note that the CCR5 ligand database without physiological pH charges gives poorer enrichments than using ligands that have been charged at pH 7. This is because the SCH derivatives require the ionizable piperidine nitrogen to be positively charged in order to interact with the binding site carboxylic oxygens.[50] Hence, the SCH family only appears in the last percentages of database screened, unless physiological pH charges are used (see the screening diversity analysis given below). With physiologically charged ligands, the AUTODOCK enrichments also show many actives at the first percentages of the database screened.

Looking at the FRED scoring functions, it can be seen that Chemgauss3 is the best scoring function for both CXCR4 and CCR5. Shapegauss also performs well for the CCR5 inhibitors. For both receptors, Screenscore and PLP perform similarly but worse than Chemgauss3, ChemScore, and

Table 3. Blind Docking Atomic Contact Statistics for the Calculated CXCR4-AMD3100 Complex[a]



distance to O− (sp3) Asp[171] (Å)	distance to O (sp2) Asp[171] (Å)	distance to O− (sp3) Asp[262] (Å)	distance to O (sp2) Asp[262] (Å)	distance to O− (sp3) Glu[288] (Å)	distance to O (sp2) Glu[288] (Å)	inside binding cavity?	computational free energy of binding, ΔG (kcal/mol)	relative position to the lowest Docked Energy conformation	computational free energy of binding (ΔG) for the lowest Docked Energy conformation (kcal/mol)
				AMD3100-CXCR4 LOOPS Modeled with CONGEN					
11.21(N26)	8.12(N23)	4.43(N11)	6.52(N8)	4.56(N23)	3.90(N26)	yes	−11.66	1	−11.66
11.63(N18)	13.74(N15)	7.41(N30)	6.98(N33)	7.82(N33)	6.94(N30)	yes	−9.93	39	−12.64
8.97(N23)	9.18(N26)	13.84(N30)	14.72(N26)	15.86(N8)	16.89(N18)	no	−8.54	61	−12.65
10.62(N33)	11.73(N23)	16.81(N15)	17.28(N11)	16.90(N15)	18.03(N11)	no	−8.93	73	−12.76
15.09(N26)	15.27(N23)	5.67(N15)	6.74(N18)	7.73(N8)	6.96(N18)	yes	−10.01	42	−10.83
			standard deviations:				1.21		0.84
				AMD3100-CXCR4 LOOPS Modeled with MODELLER					
15.73(N11)	14.94(N15)	13.70(N33)	13.30(N30)	24.23(N33)	24.32(N30)	no	−7.41	32	−10.53
13.44(N11)	10.63(N15)	14.26(N26)	14.74(N30)	24.13(N26)	24.76(N30)	no	−7.17	30	−11.83
14.69(N11)	14.42(N15)	14.30(N30)	12.42(N33)	23.48(N33)	24.33(N30)	no	−7.89	20	−11.43
17.29(N8)	15.08(N11)	11.76(N30)	13.89(N26)	24.45(N26)	22.91(N30)	no	−5.87	76	−12.15
15.94(N8)	15.04(N11)	12.58(N33)	12.78(N30)	22.76(N33)	23.62(N30)	no	−6.46	55	−10.47
			standard deviations:				0.80		0.76
				AMD3100-CXCR4 without LOOPS					
5.74(N18)	5.65(N8)	5.57(N26)	4.46(N23)	6.60(N23)	8.57(N33)	yes	−10.71	16	−11.32
6.00(N30)	5.97(N23)	3.12(N15)	2.72(N11)	8.04(N8)	6.46(N18)	yes	−8.05	50	−11.36
6.17(N30)	3.63(N33)	4.20(N18)	5.57(N8)	5.29(N8)	4.70(N11)	yes	−11.21	21	−13.17
6.87(N30)	5.00(N26)	4.28(N8)	6.46(N11)	6.90(N8)	6.16(N18)	yes	−12.14	1	−12.14
5.78(N30)	4.19(N26)	4.01(N11)	4.61(N8)	5.10(N8)	3.90(N18)	yes	−10.56	18	−11.99
			standard deviations:				1.52		0.75
			experimental ΔG AMD3100-CXCR4 (EC50) (kcal/mol):				−10.58		

[a] This table shows the distances from the carboxylic oxygens of the three key binding residues (Asp171, Asp262, and Glu288) to the eight AMD3100 nitrogens (N8, N18, N15, N11, N23, N26, N30, N33). All distances refer to the AUTODOCK pose with the lowest overall distance between the AMD nitrogens and the three key receptor residues. The AMD3100 nitrogens are charged according to physiological pH.

Table 4. Binding Pocket Docking Analysis (Conformation Closest to Known SDM Residues) for the CXCR4 Models[a]

	distance to O− (sp3) Asp^{171} (Å)	distance to O (sp2) Asp^{171} (Å)	distance to O− (sp3) Asp^{262} (Å)	distance to O (sp2) Asp^{262} (Å)	distance to O− (sp3) Glu^{288} (Å)	distance to O (sp2) Glu^{288} (Å)	computational free energy of binding, ΔG (kcal/mol)	relative position to the lowest Docked Energy conformation
				AMD3100-CXCR4 Loops Modeled with CONGEN				
	11.84(N26)	8.86(N23)	4.58(N11)	4.58(N11)	4.30(N23)	3.95(N26)	−13.28	1
	8.28(N26)	7.89(N23)	3.92(N11)	6.19(N8)	4.08(N30)	4.06(N33)	−11.76	48
	10.27(N30)	8.27(N26)	7.16(N11)	6.20(N8)	3.35(N33)	4.38(N23)	−9.70	86
	11.90(N11)	12.70(N8)	3.66(N30)	5.29(N26)	3.79(N15)	4.77(N18)	−9.27	82
	11.50(N26)	8.54(N23)	4.41(N11)	7.74(N8)	7.66(N11)	4.96(N8)	−12.65	1
average	10.76	9.25	4.75	6.00	4.64	4.42		
				standard deviation:			1.78	
				AMD3100-CXCR4 Loops Modeled with MODELLER				
	19.17(N26)	19.60(N30)	11.50(N8)	7.09(N15)	20.40(N11)	10.18(N18)	−8.74	22
	24.47(N18)	25.93(N8)	5.59(N15)	5.98(N11)	15.18(N33)	14.31(N30)	−8.53	46
	19.29(N26)	19.58(N30)	8.85(N15)	7.79(N18)	20.55(N11)	21.04(N8)	−8.77	36
	25.96(N8)	22.95(N18)	4.83(N15)	5.62(N11)	14.39(N30)	13.54(N26)	−8.23	24
	26.37(N8)	23.76(N18)	5.25(N15)	5.68(N11)	20.90(N33)	19.19(N23)	−6.71	82
average	23.05	22.36	7.20	6.43	18.28	15.65		
				standard deviation:			0.86	
				AMD3100-CXCR4 without Loops				
	5.38(N23)	5.11(N26)	5.79(N15)	3.36(N11)	4.23(N8)	3.89(N18)	−10.67	29
	6.99(N8)	3.89(N11)	6.38(N30)	3.23(N26)	5.89(N23)	7.79(N33)	−10.61	38
	6.29(N30)	3.18(N33)	5.54(N15)	3.51(N11)	4.12(N8)	3.80(N18)	−11.27	16
	6.89(N23)	3.30(N26)	4.26(N8)	3.87(N18)	5.84(N15)	3.34(N11)	−10.41	33
	6.39(N8)	3.87(N18)	5.28(N26)	3.25(N23)	6.61(N23)	10.13(N30)	−10.33	36
average	6.39	3.87	5.45	3.44	5.34	5.79		
				standard deviation:			0.37	
				experimental ΔG AMD3100-CXCR4(EC50) (kcal/mol):			−10.58	

[a] This table shows the distances from the carboxylic oxygens of the three key binding residues (Asp171, Asp262, and Glu288) to the eight AMD3100 nitrogens (N8, N18, N15, N11, N23, N26, N30, N33) measured for the AUTODOCK binding pose with the lowest overall distance between the AMD nitrogens and the key SDM residues.

Table 5. Binding Pocket Docking Analysis (Lowest Docked Energy Conformation) for the CXCR4 Models[a]

	distance to O− (sp3) Asp^{171} (Å)	distance to O (sp2) Asp^{171} (Å)	distance to O− (sp3) Asp^{262} (Å)	distance to O (sp2) Asp^{262} (Å)	distance to O− (sp3) Glu^{288} (Å)	distance to O (sp2) Glu^{288} (Å)	computational free energy of binding, ΔG (kcal/mol)
				AMD3100-CXCR4 Loops Modeled with CONGEN			
	11.84(N26)	8.86(N23)	4.58(N11)	4.58(N11)	4.30(N23)	3.95(N26)	−13.28
	10.44(N23)	9.59(N26)	4.25(N11)	7.56(N8)	7.45(N11)	4.79(N8)	−12.25
	10.54(N23)	9.62(N26)	6.44(N15)	5.61(N11)	7.68(N11)	4.94(N8)	−12.24
	11.94(N26)	8.86(N23)	6.60(N15)	5.91(N11)	7.66(N11)	5.06(N8)	−13.28
	11.50(N26)	8.54(N23)	4.41(N11)	7.74(N8)	7.66(N11)	4.96(N8)	−12.65
average	11.25	9.09	5.26	6.28	6.95	4.74	
				standard deviation:			0.52
				AMD3100-CXCR4 Loops Modeled with MODELLER			
	23.29(N18)	24.77(N15)	9.97(N33)	9.59(N23)	13.42(N18)	12.49(N8)	−10.10
	23.02(N18)	23.90(N8)	10.23(N33)	9.75(N23)	13.25(N18)	12.00(N8)	−10.36
	18.38(N11)	19.50(N8)	10.01(N33)	8.62(N30)	20.88(N30)	20.10(N26)	−9.39
	22.81(N18)	23.51(N8)	11.47(N23)	9.05(N33)	13.10(N15)	11.65(N8)	−10.01
	24.17(N33)	23.87(N23)	9.69(N15)	8.15(N18)	19.58(N15)	19.11(N18)	−8.77
average	22.33	23.11	10.27	9.03	16.05	15.07	
				standard deviation:			0.64
				AMD3100-CXCR4 without Loops			
	6.44(N30)	6.53(N26)	5.29(N11)	5.86(N8)	5.77(N8)	5.29(N18)	−12.70
	6.42(N15)	5.04(N18)	5.01(N26)	5.26(N23)	6.79(N23)	6.31(N33)	−13.32
	6.48(N15)	5.34(N18)	4.63(N26)	4.88(N23)	7.24(N23)	6.60(N33)	−13.40
	6.54(N15)	6.32(N18)	4.63(N26)	4.93(N23)	7.04(N23)	6.43(N33)	−13.31
	6.54(N15)	5.39(N18)	4.40(N26)	4.79(N23)	7.27(N23)	6.76(N33)	−13.20
average	6.48	5.72	4.79	5.14	6.82	6.28	
				standard deviation:			0.28
				experimental ΔG AMD3100-CXCR4(EC50) (kcal/mol):			−10.58

[a] This table shows the distances from the carboxylic oxygens of the three key binding residues Asp171, Asp262 and Glu288 to the eight nitrogens in AMD3100 (N8, N18, N15, N11, N23, N26, N30, N33) measured for the lowest energy AUTODOCK pose.

OEChemScore. This would seem to be because Chemgauss3 and Shapegauss are shape-based scoring functions that use smooth Gaussian functions to represent the shapes of molecules, whereas Chemgauss3 also includes a model of

Table 6. Binding Pocket Docking Analysis (Conformation Closest to Known SDM Residues and Highest GoldScore Conformation) for CXCR4[a]

AMD3100-CXCR4 LOOPS modeled with CONGEN	distance to O− (sp3) Asp^{171} (Å)	distance to O (sp2) Asp^{171} (Å)	distance to O− (sp3) Asp^{262} (Å)	distance to O− (sp2) Asp^{262} (Å)	distance to O (sp3) Glu^{288} (Å)	distance to O (sp2) Glu^{288} (Å)	relative position to the highest GoldScore conformation
Without constraints							
nearest conformation	16.86(N30)	15.80(N33)	4.97(N33)	6.24(N30)	9.24(N33)	5.72(N30)	1
highest GoldScore conformation	8.26(N26)	5.68(N30)	5.48(N11)	5.22(N23)	4.98(N23)	5.73(N8)	10
With constraints							
nearest conformation	12.31(N8)	11.20(N18)	6.75(N23)	6.30(N11)	3.82(N8)	2.60(N18)	1
highest GoldScore conformation	14.08(N30)	10.93(N33)	7.01(N15)	5.20(N11)	2.82(N11)	3.95(N23)	9

[a] This table shows the distances from the carboxylic oxygens of the three key binding residues (Asp171, Asp262, and Glu288) to the eight nitrogens in AMD3100 (N8, N18, N15, N11, N23, N26, N30, N33) calculated with and without protein hydrogen bond constraints. Each table element shows the results obtained for both the Gold pose with the lowest overall distance to the key SDM residues and for the conformation with the best GoldScore energy.

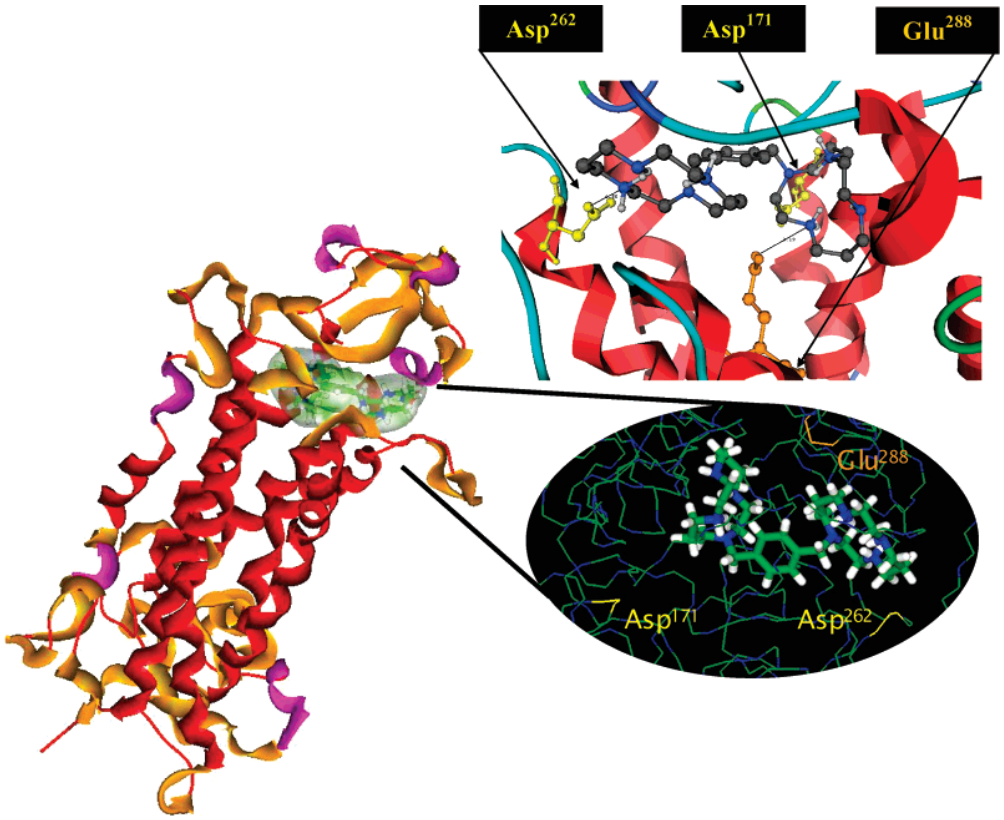


Figure 6. Close-up view of the lowest energy AMD3100-CXCR4 binding conformation. The view on the left shows AMD3100 docked within the CXCR4 pocket. The AMD3100 molecular volume is depicted using a spherical harmonic surface. The view on the right shows in detail the calculated binding conformation. In this docking prediction, two nitrogens of one cyclam ring interact with the two carboxylic oxygens of Asp262, and two nitrogens of the other cyclam ring interact with the two carboxylic oxygens of Glu288.

the molecular chemical properties. If the protein structures contain errors, as is likely with model-built structures, those scoring functions that include a chemical description of known binders might be expected to be more resilient to structural errors in the receptor. Therefore, it is perhaps not surprising that Chemgauss3 (shape plus chemistry) gives better enrichments than Shapegauss (shape-only), especially in the case of CXCR4. In this case, all screening compounds have generally similar shapes, and it is largely chemical properties that distinguish the actives from the inactives. On the other hand, CCR5 has many different families of antagonists, each with rather different shapes, so Shapegauss distinguishes them well from the decoys. The FRED Consensus Scoring improves the enrichment of FRED scoring functions used in both cases. The CXCR4 FRED Consensus Scoring enrichments are better than the CCR5 EFs in first percentages of database screened, but in both cases all the actives are found at 10% database screened.

The HEX docking function gives similar enrichments to Chemgauss3, which are better for CXCR4 than for CCR5. For the GOLD GoldScore and ChemScore scoring functions with CXCR4, ChemScore is observed to give a better enrichment than GoldScore. Consensus Rank-by-Rank scoring improves the enrichment of the GoldScore, ChemScore, and Docked Energy scoring functions, giving enrichments similar to Chemgauss3. Overall, the theoretical maximum EF for these databases are 19.9% for CXCR4 inhibitors and 14.3% for CCR5 inhibitors. Thus, the results obtained here are, in fact, rather respectable, compared to other docking-based virtual screening exercises that use modeled GPCR

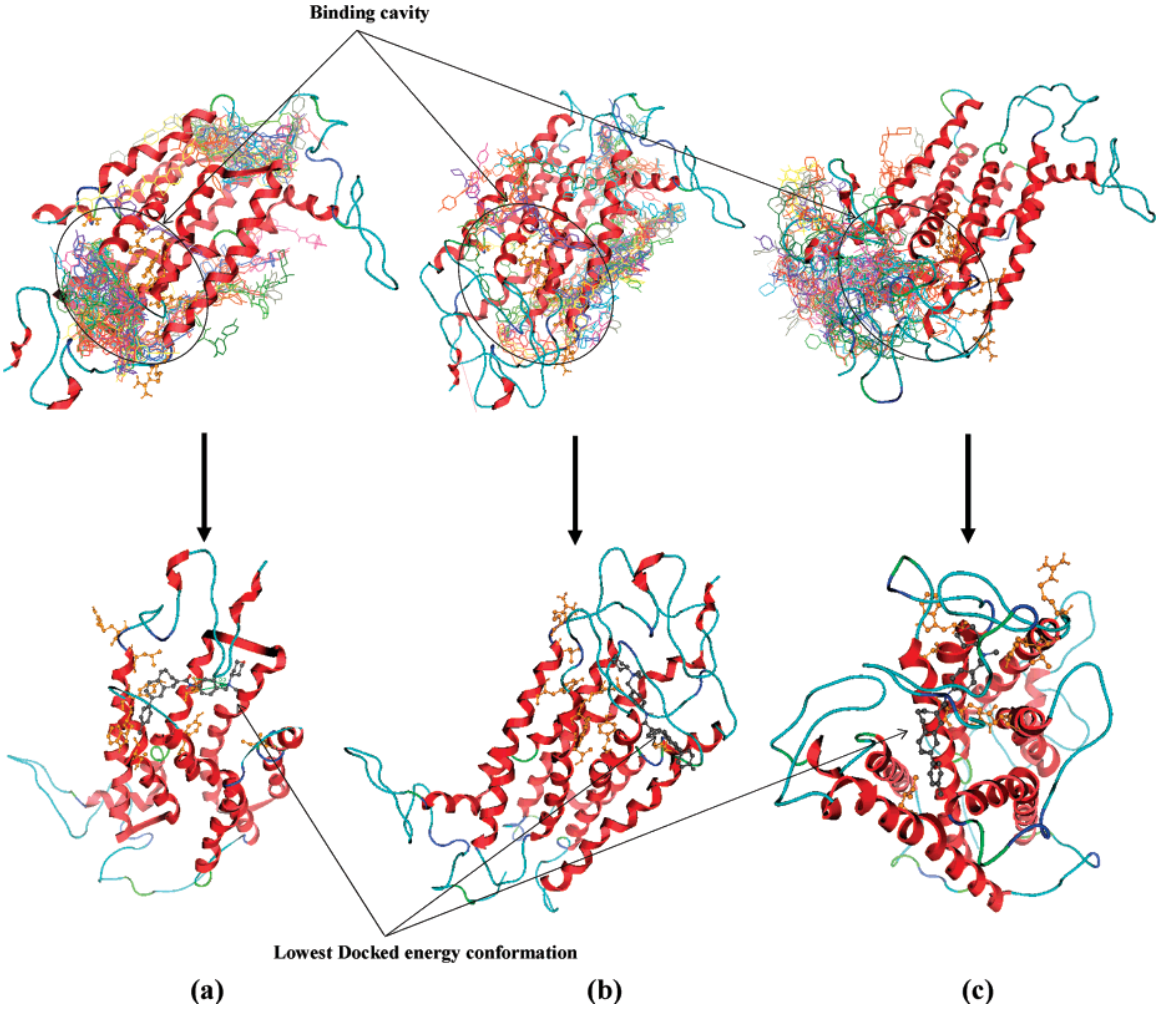


Figure 7. Blind docking results for TAK779-CCR5 obtained using AUTODOCK. Top row: the starting conformation is shown as yellow sticks, and the resulting 100 conformations are shown as colored sticks. Bottom row: the complexes are rotated to give a good view of the ligand-bound binding site. The lowest energy docked conformations are shown in each case as balls and sticks. Key binding residues are shown as orange balls and sticks. Panel a shows the NO-loops model (the lowest energy docking conformation is found within the binding pocket); panel b shows the MODELLER-loops model (no low energy conformations are found in the binding pocket, because of steric clash with EL2); and panel c shows the CONGEN-loops model (the lowest energy docking conformation is found within the binding pocket).

structures.[14,118,121]

Shape-Matching Enrichments. In many virtual screening endeavors, the crystallographic ligand from the complex is often used as the query molecule.[122] In the current study, no crystal structure information is available, so the query molecule was selected as the SDM-compatible binding conformation found from computational docking, as described previously. Figure 10 shows some example superpositions of the docked AMD3100 and TAK779 conformations from the shape-matching virtual screening procedure. In this example, ROCS, PARAFIT, and HEX all produce rather similar superpositions between the query compound and the database ligand (AMD3167 CXCR4 inhibitor and SCH417690 derivative CCR5 inhibitor). Figure 11 shows that, overall, the ROCS Combo Score gives the best EFs when using the docked AMD3100 and TAK779 conformations as queries. However, it can also be observed that the PARAFIT Shape Tanimoto gives generally better results than ROCS Shape Tanimoto and often gives results comparable to the ROCS Combo Score for both CXCR4 and CCR5 inhibitors. The HEX Shape Tanimoto functions performs well for the CXCR4 inhibitors at the first percentages of the database screened, but the EFs are considerably lower for the CCR5 inhibitors. For the CXCR4 inhibitors, the HEX Shape Tanimoto and ROCS Combo Score give EFs comparable to the theoretical maximum (19.9%) at the first percentage of database screened. For the CCR5 inhibitors, ROCS Combo Score and PARAFIT give EFs comparable to the theoretical maximum (14.3%) at the first percentage of the database screened. Moreover, for CXCR4 inhibitors, the four shape-matching scoring functions perform well at the next percentages of database screened. However, the CCR5 inhibitor EFs are generally not as good as the CXCR4 EFs, although the relative utility of the different scoring functions is similar in both cases.

The lower EFs obtained for CCR5 seem to be due to the fact that the query conformation is not able to superpose all of the CCR5 ligand families well. The query superposes well onto actives from the same scaffold (which are retrieved first) but it cannot superpose well to actives with different scaffolds. To investigate this phenomenon further, the enrichments obtained using our docked TAK779 conforma-

Table 7. Blind Docking Binding Site Analysis for the CCR5 Models[a]

		Hydrophobic Interactions (Å)										
distance to O− (sp3) Glu^{283} (Å)	distance to O (sp2) Glu^{283} (Å)	distance to Tyr^{37}	distance to Trp^{86}	distance to Tyr^{108}	distance to Leu^{33}	distance to Arg^{31}	distance to Ile^{198}	distance to Thr^{82}	inside binding cavity?	computational free energy of binding, ΔG (kcal/mol)	relative position to the lowest Docked Energy conformation	computational free energy of binding (ΔG) for the lowest Docked Energy conformation (kcal/mol)
					TAK779-CCR5 Loops Modeled with CONGEN							
7.23	5.11			C−OH(3.20)					yes	−11.56	1	−11.56
10.40	9.29								no	−12.2	7	−8.74
6.33	7.30			C-CD2(3.24)					yes	−11.3	14	−9.33
6.31	8.39								yes	−11.04	28	−12.12
7.66	8.26			C-CD2(3.15)					yes	−12.05	40	−10.90
				standard deviations:						0.49		1.45
					TAK779-CCR5 Loops Modeled with MODELLER							
10.94	12.92				C-CD1(2.87)				no	−16.40	27	−24.43
10.27	12.27				C-CD1(3.10)				no	−14.64	93	−22.21
14.43	14.39								no	−19.34	3	−19.37
10.44	12.29				C-CG(3.11)				no	−15.92	31	−19.88
10.82	12.29								no	−15.66	51	−19.95
				standard deviations:						1.77		2.13
						TAK779-CCR5 without Loops						
6.03	5.24		C-CD2(3.41)						yes	−13.64	87	−18.50
4.47	3.15	C−OH(3.25)	C-CZ3(2.89)	C−OH(2.64)					yes	−17.72	43	−18.78
4.00	3.89						C-CG2(2.88)		yes	−15.52	44	−19.82
4.17	4.06	C−OH(2.93)							yes	−16.87	35	−19.52
3.97	4.73		C−O(2.80)						yes	−14.82	49	−19.44
4.66	4.34	C−OH(3.22)		O-CE1(2.78)		C-NE(3.09)			yes	−18.47	42	−19.44
				standard deviations:						1.83		0.50
				experimental ΔG TAK779-CCR5(IC50) (kcal/mol):						−12.08		

[a] This table shows the distances from the carboxylic oxygens of Glu283 to TAK779 ammonium nitrogen and hydrophobic interactions between TAK779 and Trp86, Tyr37, Tyr108, Leu33, Arg31, Ile198 and Thr82 residues. All distances are measured for the AUTODOCK binding pose with the lowest overall distance between the TAK779 and the key SDM residues.

Table 8. Binding Pocket Docking Analysis (Conformation Closest to Known SDM Residues) for the CCR5 Models[a]

					Hydrophobic Interactions (Å)								
	distance between ammonium N+ and O− (sp3) Glu^{283} (Å)	distance between ammonium N+ and O (sp2) Glu^{283} (Å)	distance between peptide N and O− (sp3) Glu^{283} (Å)	distance between peptide N and O (sp2) Glu^{283} (Å)	distance to Tyr^{37}	distance to Trp^{86}	distance to Tyr^{108}	distance to Leu^{33}	distance to Arg^{31}	distance to Ile^{198}	distance to Thr^{82}	computational free energy of binding, ΔG (kcal/mol)	relative position to the lowest Docked Energy conformation
					TAK779-CCR5 Loops Modeled with CONGEN								
	4.13	5.46	7.65	6.72	C−OH(3.06)					C-CD1(3.37)		−11.28	14
	4.37	4.46	6.14	6.27	C−OH(2.20)		C-CE2(2.89)					−13.06	26
	5.70	3.95	7.41	6.99	C−OH(2.87)		C-CE2(3.00)					−11.71	57
	5.25	5.45	10.79	9.61			C-CD2(3.25)			C-CD1(3.17)		−13.30	60
	6.37	4.59	10.04	8.82			C-CD2(3.05)			C-CD1(3.42)		−13.79	73
average	5.16	4.78	8.41	7.68	2.71		3.05			3.32			
					standard deviation:							1.08	
					TAK779-CCR5 Loops Modeled with MODELLER								
	10.00	10.50	13.96	13.45								−18.04	42
	10.14	11.64	13.15	15.18								−18.02	17
	10.29	11.83	13.09	15.12								−14.52	48
	10.60	11.82	14.12	15.72								−15.99	69
	10.59	11.82	14.38	15.95								−15.72	78
average	10.32	11.52	13.74	15.08									
					standard deviation:							1.54	
					TAK779-CCR5 without Loops								
	4.31	3.89	7.98	8.78	C−OH(2.91)							−17.75	61
	4.31	3.65	7.99	8.87	C−OH(2.80)				C-NE(3.32)			−18.27	32
	3.92	3.53	8.07	9.12	C-CE2(2.96)		O-CD1(2.96)		C-NE(2.94)			−17.92	29
	4.68	3.54	6.78	7.30	C−OH(3.54)		C-CE1(3.12)		C-CA(3.19)			−18.99	19
	4.6	3.07	7.49	7.67	C−OH(3.59)	C-CZ3(3.06)	C−OH(3.21)					−17.09	75
average	4.36	3.54	7.66	8.35	2.89	3.06	3.10		3.15				
					standard deviation:							0.70	
					experimental ΔG TAK779-CCR5(IC50) (kcal/mol):							−12.08	

[a] This table shows the distances from the carboxylic oxygens of Glu283 to TAK779 ammonium and peptide nitrogens and hydrophobic interactions between TAK779 and Trp86, Tyr37, Tyr108, Leu33, Arg31, Ile198 and Thr82 residues. All distances refer to the AUTODOCK pose with the lowest overall distance between TAK779 and the key SDM residues.

Table 9. Binding Pocket Docking Analysis (Lowest Docked Energy Conformation) for the CCR5 Models[a]

					Hydrophobic Interactions (Å)							
	distance between ammonium N+ and O^- (sp3) Glu^{283} (Å)	distance between ammonium N+ and O (sp2) Glu^{283} (Å)	distance between peptide N and O^- (sp3) Glu^{283} (Å)	distance between peptide N and O (sp2) Glu^{283} (Å)	distance to Tyr^{37}	distance to Trp^{86}	distance to Tyr^{108}	distance to Leu^{33}	distance to Arg^{31}	distance to Ile^{198}	distance to Thr^{82}	computational free energy of binding, ΔG (kcal/mol)
					TAK779-CCR5 Loops Modeled with CONGEN							
	6.50	7.01	6.17	4.48	C−OH(3.06)					C-CD1(3.62)		−12.36
	5.44	3.68	4.63	4.86	C−OH(2.97)		C-CZ(3.25)					−13.06
	6.45	7.34	7.95	6.98	C−OH(2.86)					C-CD1(3.48)		−12.08
	6.55	4.62	7.71	7.05						C-CD1(3.51)		−13.40
	5.90	4.63	7.82	5.69	C−OH(2.30)		C-CD2(3.40)					−14.45
average	6.17	5.46	6.86	5.81	2.68		3.33			3.54		
					standard deviation:							0.94
					TAK779-CCR5 Loops Modeled with MODELLER							
	13.86	15.93						C-CG(3.55)				−18.43
	14.86	16.23							C-NH2(3.04)			−23.91
	13.59	14.95							C-NE(3.20)			−24.52
	15.43	17.50							C−N(3.07)			−19.73
	14.23	15.60							O−CB(3.18)			−26.01
average	14.39	16.04						3.55	3.12			
					standard deviation:							3.26
					TAK779-CCR5 without Loops							
	10.02	8.32	5.59	3.39	C−OH(2.88)		C-CZ(3.17)			C-CG2(3.12)		−19.97
	9.83	8.20	5.38	3.17	C−OH(2.99)		C-CZ(3.29)			C-CG2(3.34)		−19.93
	9.51	7.97	5.08	2.88	C−OH(2.89)		C-CZ(3.46)			C-CD1(3.65)		−19.75
	9.98	8.29	5.51	3.32	C−OH(2.85)		C-CE1(3.11)			C-CG2(3.07)		−20.07
	9.92	8.28	4.86	2.65	C−OH(2.98)		C-CZ(3.35)			C-CD1(3.08)		−19.83
average	9.85	8.21	5.28	3.08	2.92		3.28			3.25		
					standard deviation:							0.12
					experimental ΔG TAK779-CCR5(IC50) (kcal/mol):							−12.08

[a] This table shows the distances from the carboxylic oxygens of Glu283 to TAK779 ammonium and peptide nitrogens and hydrophobic interactions between TAK779 and Trp86, Tyr37, Tyr108, Leu33, Arg31, Ile198 and Thr82 residues in the lowest energy AUTODOCK pose.

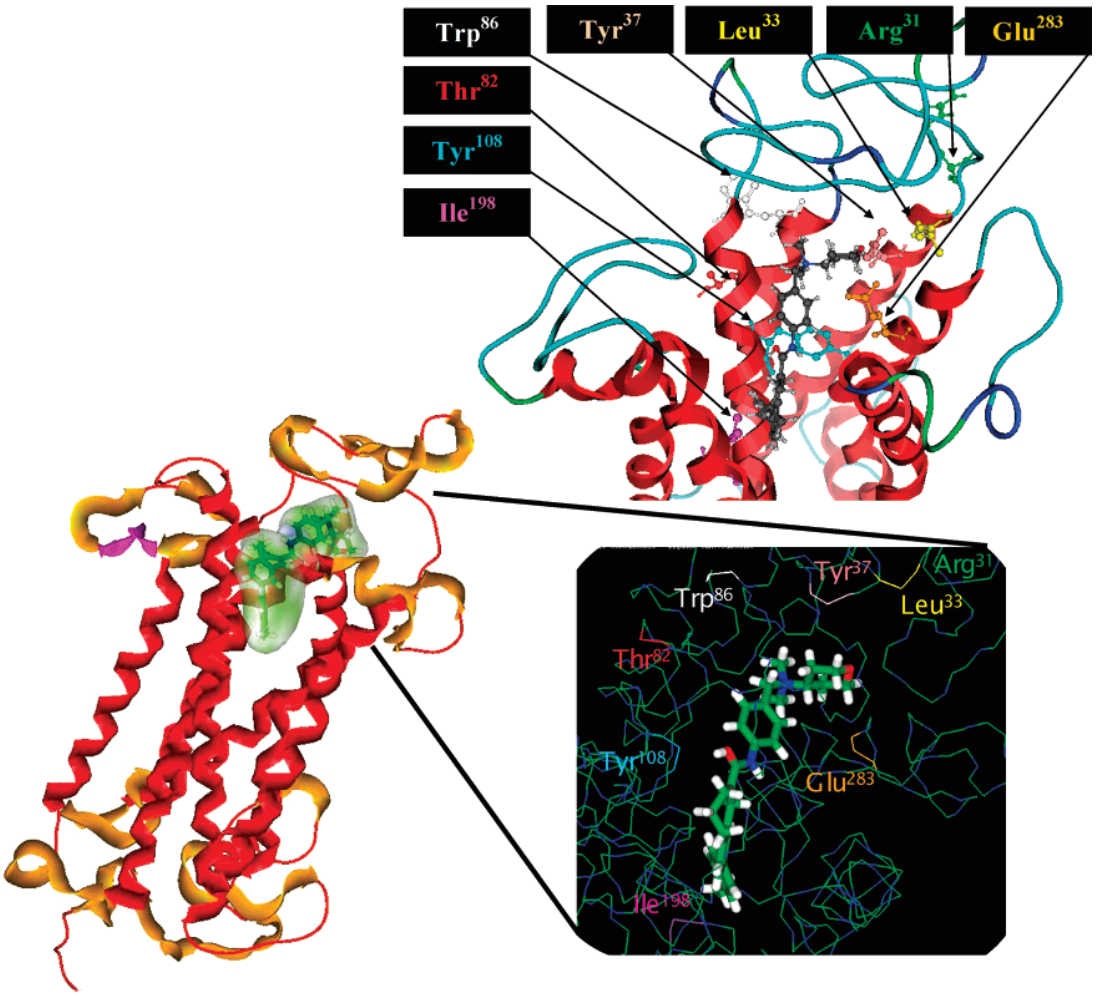


Figure 8. Close-up view of the lowest energy TAK779-CCR5 binding mode. The view on the left shows TAK779 docked within the CCR5 pocket. The TAK779 molecular volume is depicted using a spherical harmonic surface. The view on the right shows in detail the calculated binding mode conformation. The TAK779 benzyl pyran ammonium group interacts with helices TM1, TM2, and TM7 with a near contact between ammonium nitrogen of TAK779 and Glu283 and a hydrophobic interaction with Tyr37, whereas the methylphenyl-benzocycloheptenyl moiety is buried in the TM barrel and makes hydrophobic interactions with Tyr108 and Ile198.

tion (calculated as described previously) were compared to those calculated for the ligand conformation used by Fano et al.[116] (minimized using the AMBER/MM2 force field and docked using the QXP DYNDOCK module). The influence of different conformations of the query, as well as different conformations of the database compounds, using ROCS Shape Tanimoto and ComboScore functions, was also analyzed. These results are shown in Figure 12. Regarding the two ligand conformations, the conformation of Fano et al. gives slightly better EFs with the PARAFIT Shape Tanimoto, whereas the ROCS Shape Tanimoto, ROCS Combo and HEX Shape Tanimoto scores give slightly better EFs with our TAK779 conformation at the first percentages of database screened. At the last percentages of database screened, the Fano et al. query seems to discriminate more actives from decoys than our query. Nonetheless, both queries give generally similar overall EFs. Regarding the diversity of query conformations, the ROCS Shape Tanimoto and ROCS Combo functions both perform similarly with one or ten query conformations. In both cases the EFs improve a little in the first percentages of the screened database but subsequently decline at the lower percentages. Moreover, using 10 query conformations and 10 conformations of every database compound does not substantially improve the EFs compared to using only one query conformation. These results suggest that no single active is able to superpose the remaining inhibitor families well. Consequently, this implies that there is probably more than one binding mode within the CCR5 pocket.

Comparison of Docking and Shape-Matching Results. Figure 13 shows a comparison of the docking-based and ligand-based enrichments obtained for the CXCR4 and CCR5 inhibitors. This figure shows that the similarity-based functions give much better enrichments for CXCR4 than for CCR5, because of the difficulty in finding good CCR5 inhibitor conformations as mentioned previously. For shape-only comparisons, PARAFIT generally gives better EFs than the ROCS Shape Tanimoto and often gives comparable EFs to the ROCS Combo Score in both cases. Results for CCR5 show that the ROCS Combo Score and PARAFIT Shape Tanimoto give similar but somewhat modest enrichments, being comparable to the FRED Chemgauss3, Consensus, Shapegauss, and AUTODOCK Docked Energy functions at the first percentages of screened database. Of the docking algorithms, the highest enrichments at the first percentages of database screened are found using FRED Consensus and

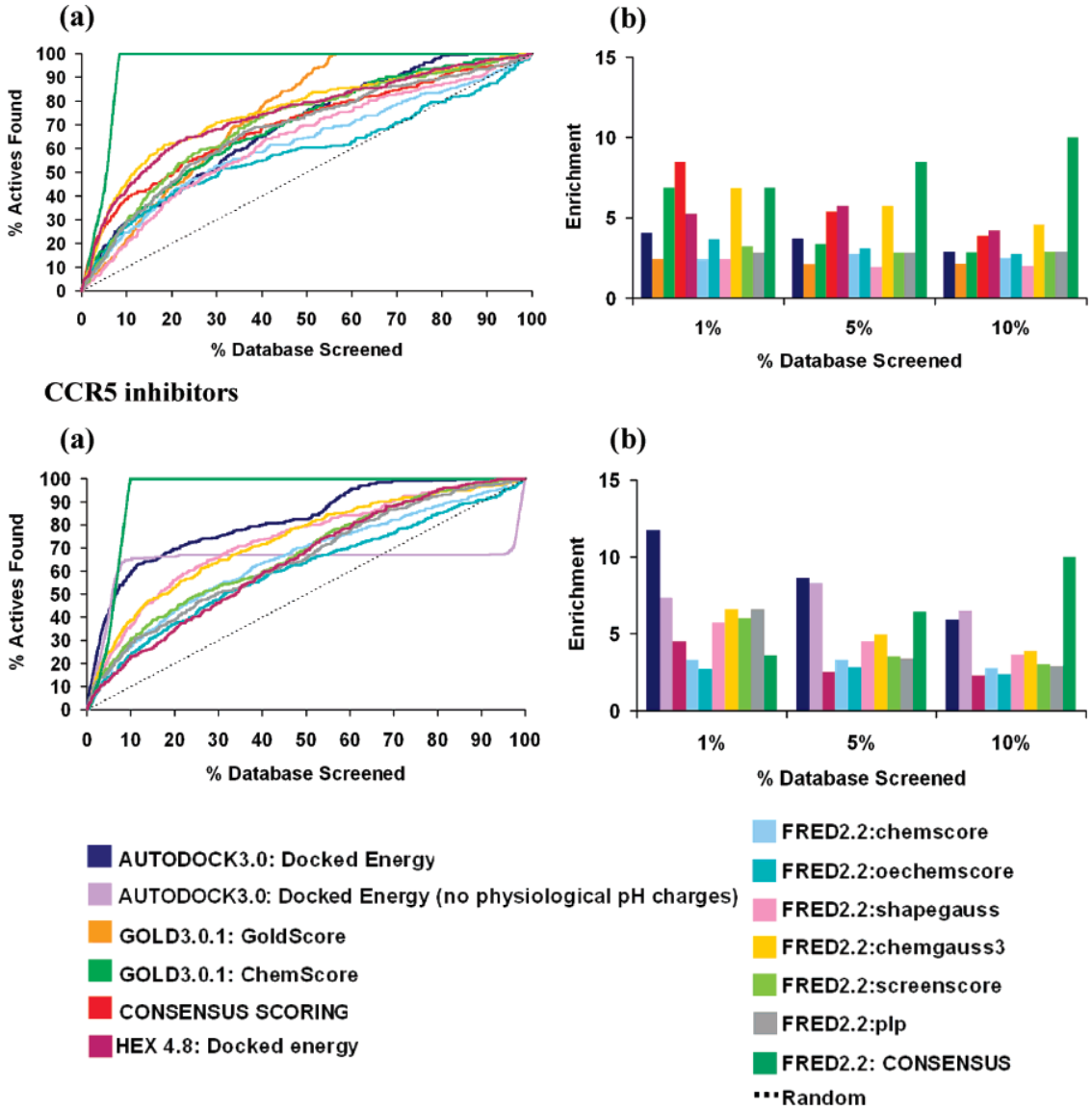


Figure 9. CXCR4 and CCR5 docking-based enrichments: (a) enrichment results for several docking protocols applied to CXCR4 and CCR5 receptors (the dotted black line represents the values expected if actives were selected at random), and (b) enrichment factor for actives found within the top-ranking 1%, 5%, and 10% fractions of the CXCR4 and CCR5 screened databases.

Docked Energy. The corresponding results for CXCR4 show that Consensus Scoring, Chemgauss3, and FRED Consensus are the best of the docking methods; however, they are nonetheless worse than the ligand-based methods. It is also worth mentioning that, for both receptors, the Shapegauss docking scoring function performs similarly, although with lower EFs, to the ROCS Shape Tanimoto, and the Chemgauss3 docking function performs similarly to the ROCS Combo Score.

Screening Diversity Analysis. The ability of the docking-based and ligand-based approaches to retrieve a diverse scaffold pool that might facilitate the identification of novel lead structures was assessed by determining the number of actives found for each scaffold class at several percentages of the ranked database. Figure 14 summarizes the retrieval rates for the 13 families of CCR5 and 4 families of CXCR4 ligands, as listed in Table 1. This figure shows that, at 5% screened database, all of the CXCR4 scaffolds are found by the ROCS Combo, FRED Consensus, and FRED Chemgauss3 scoring functions, whereas for CCR5, only AUTODOCK Docked Energy and FRED Consensus found them all. At 10% screened database, all of the CXCR4 scaffolds are found by the various docking and shape-matching scoring functions. However, for CCR5, the docking scoring functions found all of the scaffolds at 20% screened database, whereas the shape-matching functions do not recognize 4 of the 13 families (i.e., AMD derivatives, piperidine, aminopiperidine, *N*,*N*′-diphenylurea and 5-oxopyrrolidine-3-carboxamide scaffolds).

Results for CCR5 show that, at 1% database screened, the first scaffolds retrieved are the TAK derivatives, some SCH derivatives, some 1-phenyl-1,3-propanodiamine, and some anilide piperidine N-oxide compounds. Considering the total number of compounds of every family in the database, it can be observed that the SCH derivatives, TAK derivatives, 1-phenyl-1,3-propanodiamines, aminopiperidines, and anilide piperidine N-oxide families have more representatives, so it is natural to find more actives from these families within the first percentages of the database. Moreover, results for CCR5 show that using AUTODOCK with compounds not

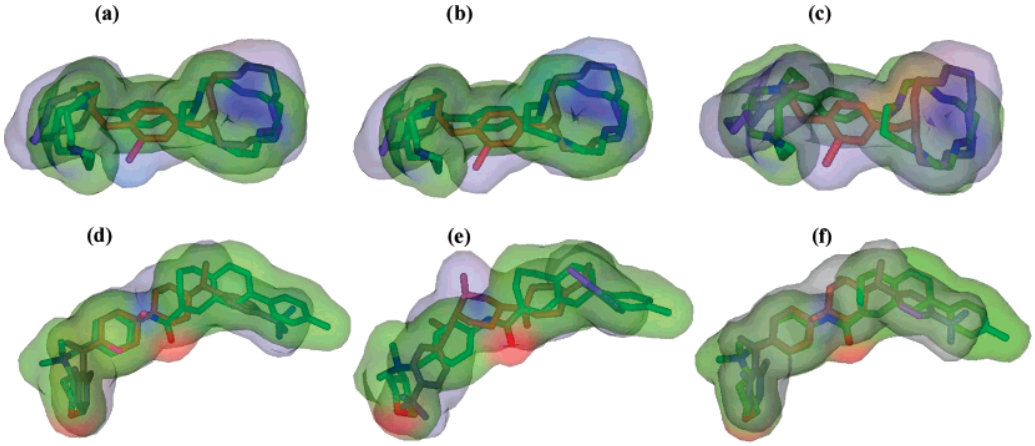


Figure 10. Example superpositions from the shape-matching virtual screening procedure. Each database compound is shown in blue/red, and the color of the query molecule is dependent on the atom type (blue, nitrogen; green, carbon; red, oxygen). Images a, b, and c show a database compound (AMD3167) superposed to AMD3100 with PARAFIT, ROCS, and HEX, respectively; images d, e, and f show a database compound (SCH417690 derivative) superposed to TAK779 using the same software, respectively. All images are drawn using HEX SH surface overlays.

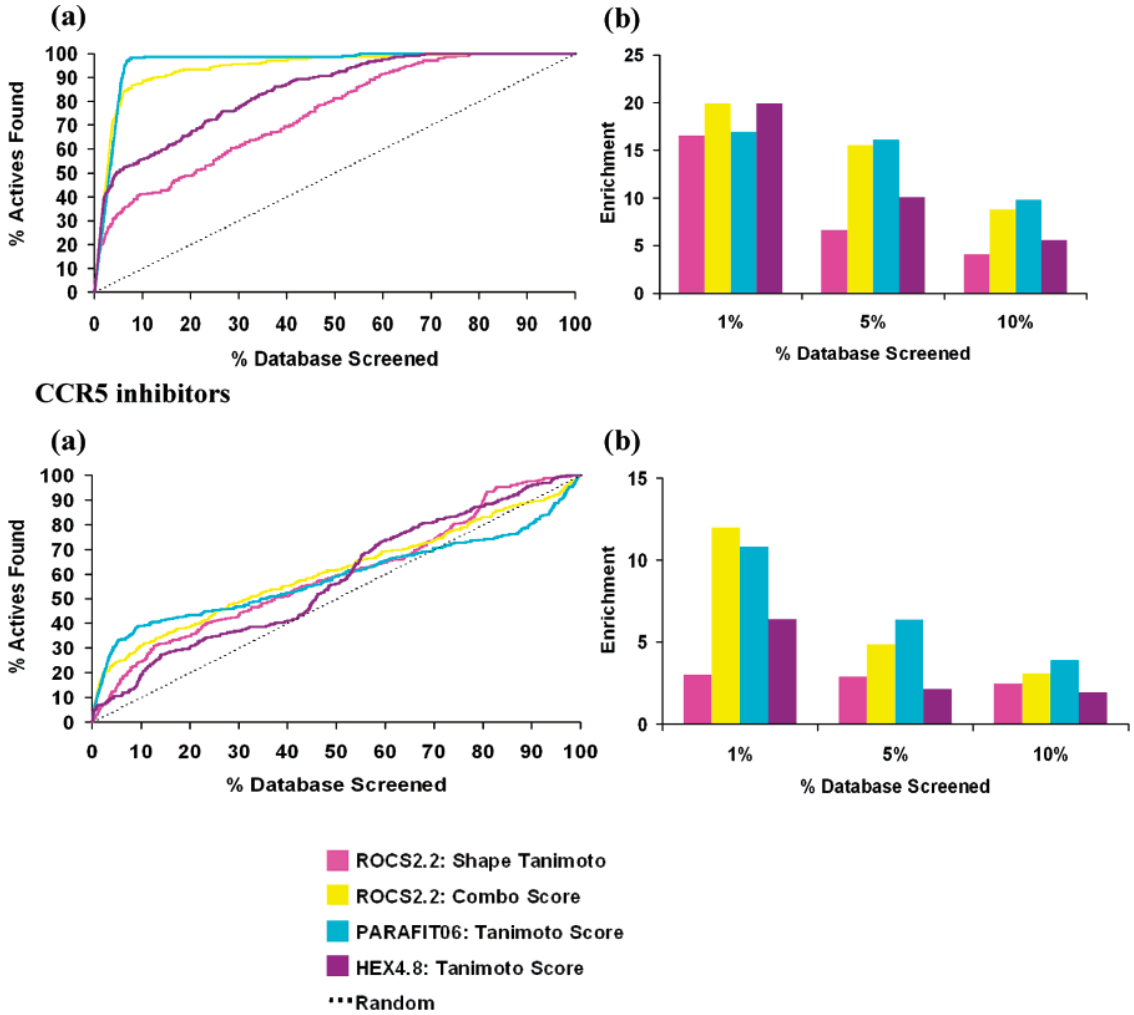


Figure 11. CXCR4 and CCR5 shape-matching-based enrichments: (a) enrichment curves obtained using several shape-matching protocols with the CXCR4 and CCR5 inhibitors (the dotted line represents the expected enrichment if actives were selected at random), and (b) enrichment curves for actives found within the top-ranking 1%, 5%, and 10% of compounds of the CXCR4 and CCR5 screened databases.

charged at physiological pH gives poorer enrichments than with charged compounds (i.e., by not finding, e.g., SCH, AMD, or diketopiperazine families in the first percentages of database screened). Conversely, the results for CXCR4 show that the shape-matching functions retrieve most of the scaffolds within the first percentages of database screened. This is consistent with the high enrichments found for the CXCR4 inhibitors, confirming that ligand-based screening

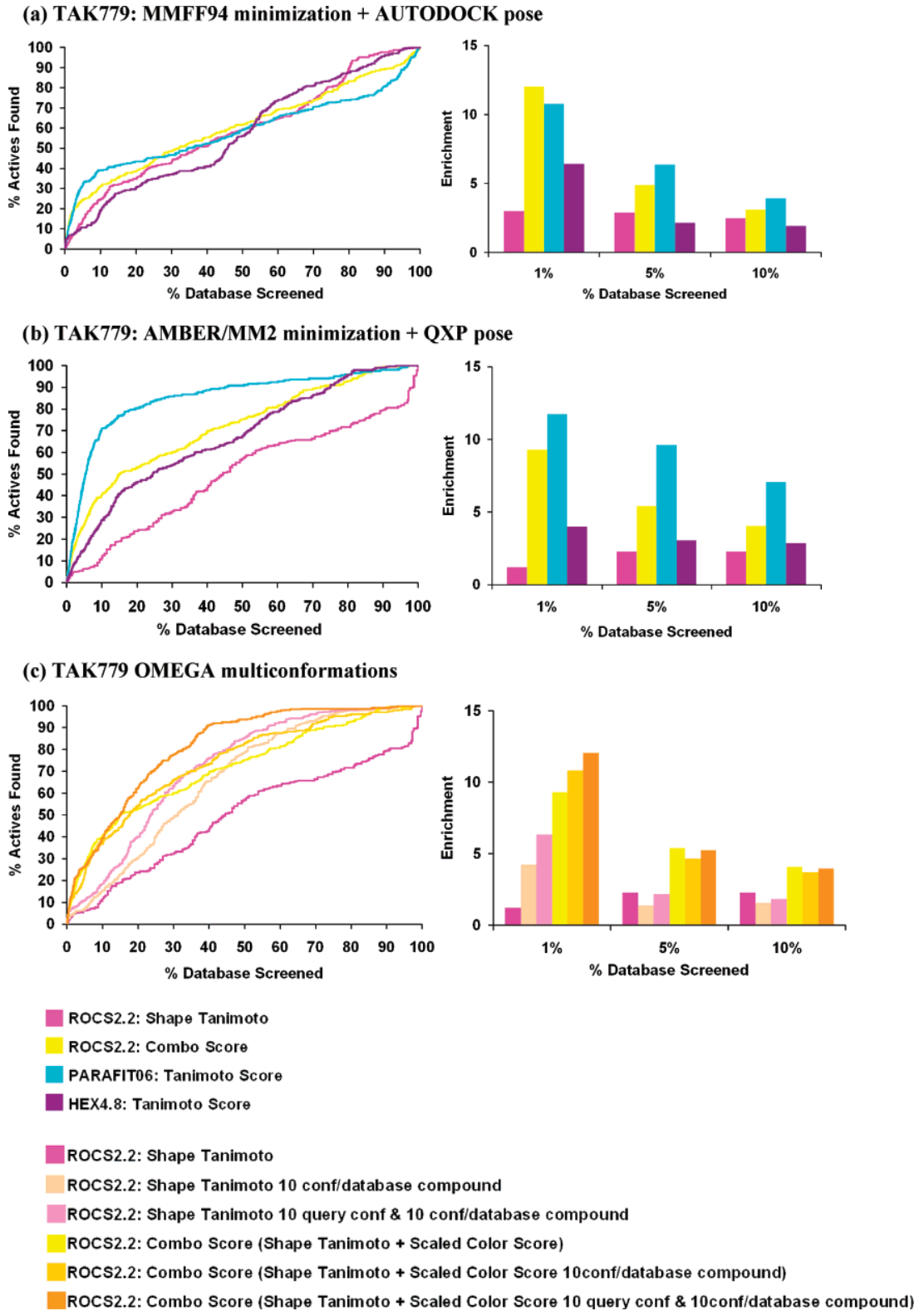


Figure 12. Comparison of enrichments for different TAK779 query conformations. Enrichment curve and enrichment factor plots obtained in shape-matching virtual screening using as query: (a) TAK779 minimized with MMFF94 force field and docked with AUTODOCK, (b) TAK779 minimized with AMBER/MM2 force field and docked with QXP, and (c) multiple conformations of TAK779 calculated by OMEGA.

is superior to docking-based approaches in this case. Nonetheless, the docking-based approaches also find diverse scaffolds for CXCR4. Indeed, the Rank-by-Rank approach identifies the same scaffolds as the Docked Energy, GoldScore, and ChemScore scoring functions. FRED Consensus identifies all CXCR4 compounds at 10% database screened and finds all CCR5 ligands at 20% database screened. Finally, it is also worth noting that, for both receptors, the

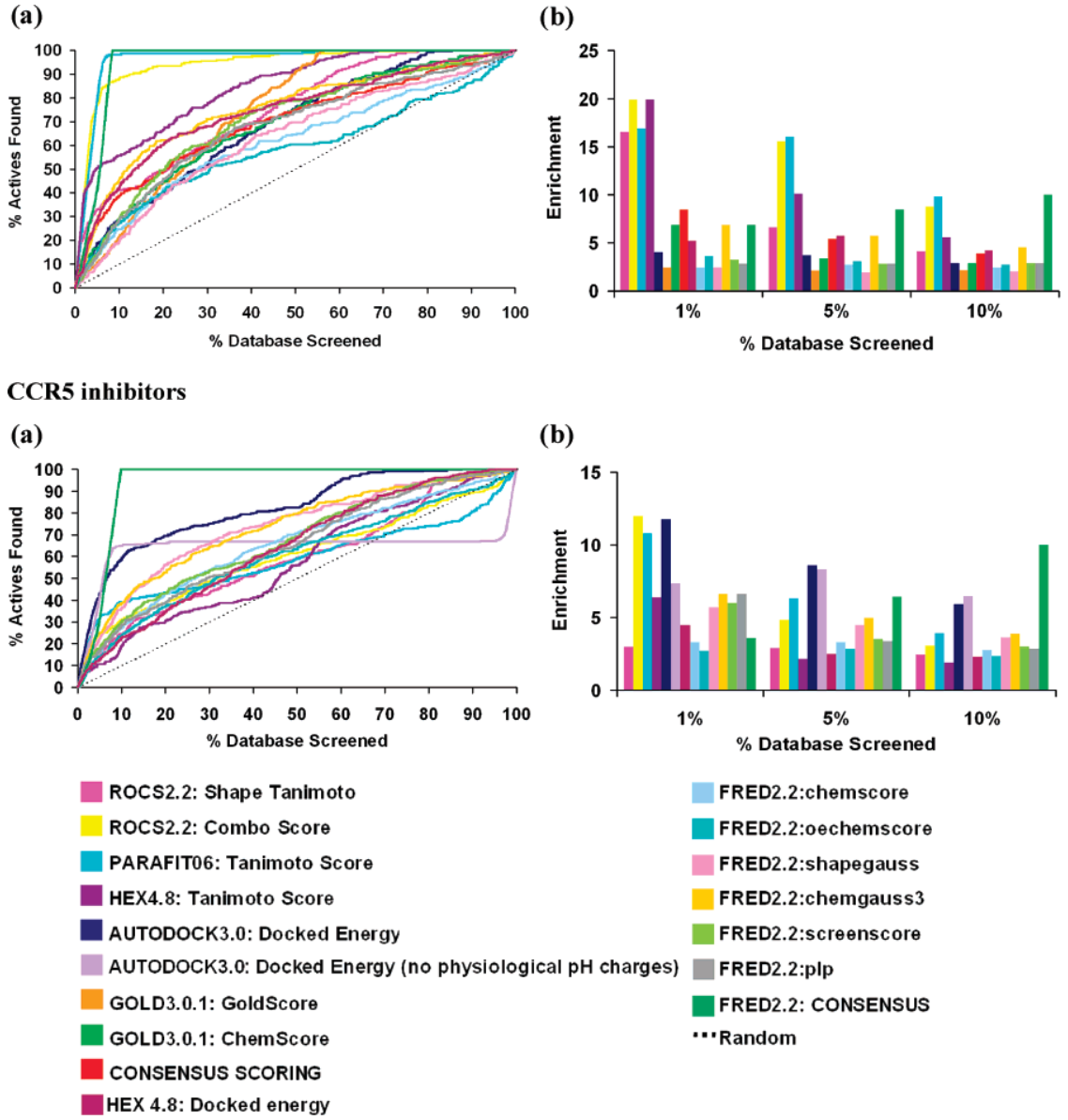


Figure 13. Comparison of docking and shape-matching enrichments for CXCR4 and CCR5 receptors: (a) enrichment curves obtained using several docking and shape-matching protocols applied to CXCR4 and CCR5 receptors (the dotted black line represents the values expected if actives were selected at random), and (b) enrichment factor for actives found within the top-ranking 1%, 5%, and 10% compounds of the CXCR4 and CCR5 screened databases.

shape-matching methods always first find those compounds with scaffolds from the same family as the query.

DISCUSSION

The results of this study indicate that our CXCR4 receptor model and the supposed binding mode for active molecules are broadly correct, inasmuch as our enrichment plots exhibit very good recognition of the known actives. Hence, it is now feasible to use our receptor model and database to perform prospective virtual screening to find new active CXCR4 antagonist compounds.[123,124] Overall, the enrichment results for this receptor show that ligand-based shape-matching approaches provide better performance than structure-based docking tools. However, the enrichments obtained for CCR5 are not as good as those for CXCR4. The CCR5 co-receptor seems to have a larger binding pocket than CXCR4 and, for this reason, it is difficult for docking algorithms to locate feasible binding modes of the known actives. In addition, MD analyses of the EL2 of CCR5 suggest that this region is highly flexible and may serve as a flexible lid or latch that constrains the ligand within the adjoining pocket. A comparison of the docking results obtained with our CCR5 co-receptor models with those of Fano et al.[116] support the notion that the EL2 conformation has a critical effect on ligand recognition. Although the model of Fano et al. recognizes the TAK derivatives well, it cannot accommodate some of the other known actives, whereas our model has a more open pocket which accepts many more of the actives but also more inactives. The available SDM data for CCR5 shows that at least eight residues (Glu283, Tyr108, Tyr37, Trp86, Leu33, Ile198, Arg31, Thr82) are critical for antagonist binding. In our model-built receptor binding site these residues form a pocket between the EL loops and the TM framework region. However, using these residues to constrain

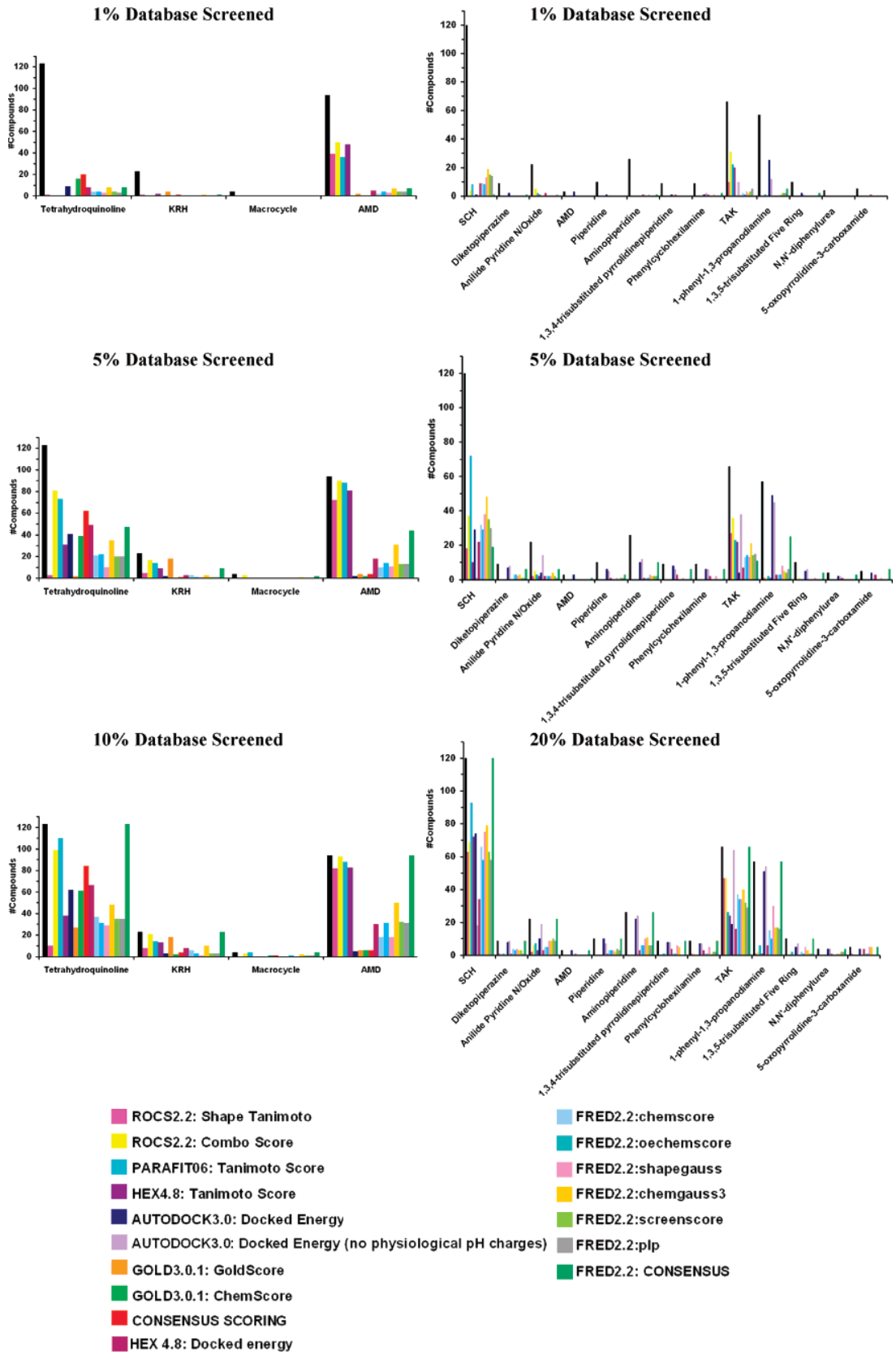


Figure 14. Scaffold diversity retrieval analysis for CXCR4 and CCR5 inhibitors. Left column shows a scaffold diversity retrieval analysis for CXCR4 inhibitors, showing the number of actives for each of 4 scaffold structures (tetrahydroquinolinamines, KRH derivatives, macrocycles, and AMD derivatives) found at 1%, 5%, and 10% of the ranked database. Right column shows a scaffold diversity retrieval analysis for CCR5 inhibitors showing the number of actives for each of 13 scaffold structures (SCH derivatives, diketopiperazines, anilide piperidine N-oxides, AMD derivatives, 4-piperidines, 4-aminopiperidines, 1,3,4-trisubstituted pyrrolidinepiperidines, phenylcyclohexilamines, TAK derivatives, 1-phenyl-1,3-propanodiamines, 1,3,5-trisubstituted pentacyclics, *N*,*N*′-diphenylureas, and 5-oxopyrrolidine-3-carboxamides) found at 1%, 5% and 20% of the ranked database.

the docking search space does not improve the results substantially. This indicates that inaccuracies exist in both receptor models, and that ligand-based techniques are needed to help identify new inhibitors for this system.

The general problem of how to select suitable query conformations is one of the confounding questions in ligand-based shape-matching virtual screening.[125] Our ligand-based shape-matching results show that it is difficult to obtain satisfactory superpositions of all of the known CCR5 antagonists. None of the multiple TAK779 query conformations individually superpose well onto all of the known actives. Nonetheless, multiple actives are known for CCR5, and all of these compounds must ultimately fit physically into the same pocket. On the other hand, the key SDM residues are spatially well-distributed around the CCR5 pocket, which suggests a large binding region. One way to unify these observations is to hypothesize that there may, in fact, be more than one binding region within the CCR5 pocket and that the actives distribute around this pocket in more than one cluster. This would explain the difficulty of finding a satisfactory global superposition of all of the known binders. This multilple-binding-region hypothesis is also supported by Castonguay et al.,[126] who determined that the binding site for the 2-aryl-4-(piperidin-1-yl)butanamine and 1,3,4-trisubstituted pyrrolidine inhibitors is located in a region similar to that proposed for other GPCR small molecule binding sites and partially overlaps the proposed TAK-779 binding site. Kellenberg et al.[14] also have cited experimental evidence for a delocalized CCR5 antagonist binding site. We are working to investigate the distributed binding site hypothesis by extending the SH-base shape-matching algorithm, to be able to identify clusters of structural scaffolds from a large set of known actives.

Overall, this study shows that homology models of the CXCR4 and CCR5 receptors built from the antagonist-bound rhodopsin template have proven to be suitable for structure-based virtual screening of HIV entry inhibitors. However, the quality of the receptor models, especially in the modeled loop regions, is critical for recognition of known binders. Our results show that if a good receptor model can be built, as in the case of CXCR4, then good enrichments can be obtained. On the other hand, homology-built protein structures will inevitably contain some errors or inaccuracies. Our ligand-based screening results show that if a set of known actives are available, then ligand shape-matching searches give better enrichments than structure-based docking, especially for CXCR4.

CONCLUSION

Molecular models of the human CXCR4 and CCR5 co-receptors were homology-built from the bovine rhodopsin X-ray crystal structure. The resulting 3D structures have good PROCHECK stereochemical statistics, and both were validated by blind docking of high-affinity antagonists. The docking modes obtained with these ligands are compatible with the available SDM data on key ligand binding residues. A large database of CXCR4/CCR5 inhibitors and similar presumed inactive compounds was compiled from the literature to perform retrospective virtual screening of antagonists against these co-receptors. This database was used to compare docking-based and ligand-based virtual screening approaches. The enrichment and diversity results obtained show that ligand-based searches are superior to docking-based approaches, especially in the case of the CXCR4 inhibitors. The virtual screening enrichments found for CCR5 were generally lower than for CXCR4 for both docking-based and ligand-based protocols. Analysis of our results suggests that there is probably more than one binding region within the CCR5 pocket and that the known antagonists distribute over this region in more than one cluster. The SH superposition approach is being extended to identify multiple scaffold superpositions to explore this hypothesis. Developing more-sophisticated 3D ligand-based screening approaches will help develop a better molecular model of CCR5-antagonist binding and will be useful for prospective virtual screening of novel scaffolds for the rational design of HIV entry blockers.

ACKNOWLEDGMENT

The authors are grateful to OpenEye Scientific Software, Inc., for providing an Academic Licence for ROCS, and to Cepos Insilico, Ltd., for providing a pre-release version of PARASURF. VIPN thanks the Generalitat de Catalunya—DURSI for a grant within the Formació de Personal Investigador (2008FI) Program. This work was supported by The TV3 Marathon Foundation (AIDS-2001) promoted by the Catalan Radio and Television Corporation (Corporació Catalana de Ràdio i Televisió, CCRTV) and the Programa Nacional de Biomedicina (Ministerio de Educación y Ciencia, SAF2007-63622-C02-01).

REFERENCES AND NOTES

(1) World Health Organisation AIDS Update 2006, http://www.who.int/hiv/en/ (accessed Dec. 23, 2007).
(2) De Clercq, E. Emerging anti-HIV drugs. *Expert Opin. Emerg. Drugs* **2005**, *10*, 241−274.
(3) De Clercq, E. Anti-HIV chemotherapy: current state of the art. *Med. Chem. Res.* **2004**, *13*, 439−478.
(4) Jiang, S.; Lin, K.; Strick, N.; Neurath, A. R. Inhibition of HIV-1 infection by a fusion domain binding peptide from the HIV-1 envelope glycoprotein GP41. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 533−538.
(5) Kadow, J.; Wang, H. G.; Lin, P. F. Small-molecule HIV-1 gp120 inhibitors to prevent HIV-1 entry: an emerging opportunity for drug development. *Curr. Opin. Invest. Drugs* **2006**, *7*, 721−726.
(6) Berger, E. A.; Murphy, P. M.; Farber, J. M. Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annu. Rev. Immunol.* **1999**, *17*, 657−700.
(7) De Clercq, E. New antiviral agents in preclinical or clinical development. *Adv. Antiviral Drug Des.* **2004**, *4*, 1−62.
(8) De Clercq, E. New Anti-HIV Agents and Targets. *Med. Res. Rev.* **2002**, *22*, 531−565.
(9) Bean, P. New Drugs Targets for HIV. *Clin. Infect. Dis.* **2005**, *41*, 96−100.
(10) Markovic, I.; Clouse, K. A. Recent advances in understanding the molecular mechanisms of HIV-1 entry and fusion: revisiting current targets and considering new options for therapeutic intervention. *Curr. HIV Res.* **2004**, *2*, 223−234.
(11) Kazmierski, W. M.; Peckman, J. P.; Duan, M.; Kenakin, T. P.; Jenkinson, S.; Gudmundsson, K. S.; Piscitelli, S. C.; Feldman, P. L. Recent Progress in the Discovery of New CCR5 and CXCR4 Chemokine Receptor Antagonists as Inhibitors of HIV-1 Entry. Part 2. *Curr. Med. Chem.—Anti-Infect. Agents* **2005**, *4*, 133−152.
(12) Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P. A.; Markopoulos, J.; Igglessi-Markopoulou, O. Investigation of substituent effect of 1-(3,3-diphenylpropyl)-piperidinyl phenylacetamides on CCR5 binding affinity using QSAR and virtual screening techniques. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 83−95.
(13) Aher, Y. D.; Agrawal, A.; Bharatam, P. V.; Garg, P. 3D-QSAR studies of substituted 1-(3, 3-diphenylpropyl)-piperidinyl amides and ureas as CCR5 receptor antagonists. *J. Mol. Model.* **2007**, *13*, 519−529.

(14) Kellenberger, E.; Springael, J.-Y.; Parmentier, M.; Hachet-Haas, M.; Galzi, J.-L.; Rognan, D. Identification of Nonpeptide CCR5 Receptor Agonists by Structure-based Virtual Screening. *J. Med. Chem.* **2007**, *50*, 1294−1303.
(15) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**, *289*, 739−745.
(16) Vaidehi, N.; Floriano, W. B.; Trabanino, R.; Hall, S. E.; Freddolino, P.; Choi, E. J.; Zamanakos, G.; A. Goddard, W., III. Prediction of structure and function of G protein-coupled receptors. *Proc. Natl. Acad. Sci., U.S.A.* **2002**, *99*, 12622−12627.
(17) Pierson, T. C.; Doms, R. W. HIV-1 entry inhibitors: new targets, novel therapies. *Immunol. Lett.* **2003**, *85*, 113−118.
(18) Kazmiersky, W.; Bifulco, N.; Yang, H.; Boone, L.; DeAnda, F.; Watson, C.; Kenakin, T. Recent progress in discovery of small-molecule CCR5 chemokine receptor ligand as HIV-1 inhibitors. *Bioorg. Med. Chem.* **2003**, *11*, 2663−2676.
(19) Kilby, J. M.; Eron, J. J. Novel therapies based on mechanisms of HIV-1 cell entry. *N. Engl. J. Med.* **2003**, *348*, 2228−2238.
(20) Starr-Spirres, L. D.; Collman, R. G. HIV-1 entry and entry inhibitors as therapeutic agents. *Clin. Lab. Med.* **2002**, *22*, 681−701.
(21) De Clercq, E. The bicyclam AMD3100 story. *Nat. Rev. Drug. Discovery* **2003**, *2*, 581−587.
(22) Tusnády, G. E.; Simon, I. Principles Governing Amino Acid Composition of Integral Membrane. Proteins: Application to Topology Prediction. *J. Mol. Biol.* **1998**, *283*, 489−506.
(23) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *J. Mol. Biol.* **2001**, *305*, 567−580.
(24) Jones, D. T.; Taylor, R.; Thornton, J. M. A Model Recognition Approach to the Prediction of All-Helical Membrane Protein Structure and Topology. *Biochemistry* **1994**, *33*, 3038−3049.
(25) Gerlach, L. O.; Skerlj, R. T.; Bridger, G. J.; Schwartz, T. W. Molecular Interaction of Cyclam and Bicyclam Non-peptide Antagonists with the CXCR4 Chemokine Receptor. *J. Biol. Chem.* **2001**, *276*, 14154−14160.
(26) Paterlini, M. G. Structure Modeling of the Chemokine Receptor CCR5: Implications for Ligand Binding and Selectivity. *Biophys. J.* **2002**, *83*, 3012−3031.
(27) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restrains. *J. Mol. Biol.* **1993**, *234*, 779−815.
(28) *MOE (Molecular Operating Environment*), 2006.08 Release; Chemical Computing Group, Inc.: Montreal, Canada, 2004.
(29) Im, D.-S. Orphan G Protein-Coupled Receptors and Beyond. *Jpn. J. Pharmacol.* **2002**, *90*, 101−106.
(30) Fiser, A.; Sali, A. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753−1773.
(31) Bruccoleri, R. E. Application of Systematic Conformational Search to Protein Modeling. *Mol. Simul.* **1993**, *10*, 151−174.
(32) Bruccolieri, R. E.; Karplus, M. Chain closure with bond angle variations. *Macromolocules* **1985**, *18*, 2767−2773.
(33) Bruccolieri, R. E.; Karplus, M. Prediction of the Folding of Short Polypeptide Segments by Uniform Conformatinal Sampling. *Biopolymers* **1987**, *26*, 137−168.
(34) Brooks, B. R.; Bruccolieri, R. E.; Olafson, B. D.; States, D. J.; Swamminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(35) Li, H.; Tejero, R.; Monleon, D.; Bassolino-Klimas, D.; Abate-Shen, C.; Bruccolieri, R. E.; Montelione, G. T.; Homology modeling using simulated annealing of restrained molecular dynamics and conformational search with CONGEN: aplication in predicting the three-dimensional structure of murine homeodomain Msx-1. *Protein Sci.* **1997**, *6*, 956−970.
(36) Bruccolieri, R. E.; Haber, E.; Novotny, J. Structure of antibody hypervariable loops reproduced by conformational search algorithm. *Nature* **1988**, *335*, 564−568.
(37) Laskowski, R. A.; McArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(38) Huang, X.; Shen, J.; Cui, M.; Shen, L.; Luo, X.; Ling, K.; Pei, G.; Jiang, H.; Chen, K. Molecular Dynamics Simulations on SDF-1α: Binding with CXCR4 Receptor. *Biophys. J.* **2003**, *84*, 171−184.
(39) Cornell, W. D.; Cieplack, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, Jr.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(40) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Hart, W.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.
(41) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 609−623.
(42) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access To Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(43) Zhang, W. B.; Navenot, J. M.; Haribabu, B.; Tamamura, H.; Hiramatu, K.; Omagari, A.; Pei, G.; Manfredi, J. P.; Fujii, N.; Broach, J. R.; Peiper, S. C. A. Point Mutation That Confers Constitutive Activity to CXCR4 Reveals That T140 Is an Inverse Agonist and That AMD3100 and ALX40−4C Are Weak Partial Agonists. *J. Biol. Chem*. **2002**, *277*, 24515−24521.
(44) Brelot, A.; Heveker, N.; Montes, M.; Alizon, M. Identification of Residues of CXCR4 Critical for Human Immunodeficiency Virus Coreceptor and Chemokine Receptor Activities. *J. Biol. Chem.* **2000**, *275*, 23736−23744.
(45) Donzella, G. A.; Schols, D.; Lin, S. W.; Este, J. A.; Nagashima, K. A.; Maddon, P. J.; Allaway, G. P.; Sakmar, T. P.; Henson, G.; De Clercq, E.; Moore, J. P. AMD3100, a small molecule inhibitor of HIV-1 entry via the CXCR4 co-receptor. *Nat. Med.* **1998**, *4*, 72−77.
(46) Hatse, S.; Princen, K.; Gerlach, L. O.; Bridger, G.; Henson, G.; Clercq, E.; Schwartz T. W.; Schols, D. Mutation of Asp171 and Asp262 of the chemokine receptor CXCR4 impairs its coreceptor function for human immunodeficiency virus-1 entry and abrogates the antagonistic activity of AMD3100. *Mol. Pharmacol.* **2001**, *60*, 164−173.
(47) Hatse, S.; Princen, K.; Vermeire, K.; Gerlach, L.-O.; Rosenkilde, M. M.; Schwartz, T. W.; Bridger, G.; De Clercq, E.; Schols, D.; Mutations at the CXCR4 interaction sites for AMD3100 influence anti-CXCR4 antibody binding and HIV-1 entry. *FEBS Lett.* **2003**, *546*, 300−306.
(48) Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R. Site-directed mutagenesis by overlap extension using polymerase chain reaction. *Gene* **1989**, *77*, 51−59.
(49) Rosenkilde, M. M.; Gerlach, L. O.; Jacobsen, J. S.; Skerlj, R. T.; Bridger, G. J.; Schwartz, T. W. Molecular Mechanism of AMD3100 Antagonism in the CXCR4 Receptor. *J. Biol. Chem.* **2004**, *279*, 3033−3041.
(50) Seibert, C.; Ying, W.; Gavrilov, S.; Tsamis, F.; Kuhmann, S. E.; Palani, A.; Tagat, J. R.; Clader, J. W.; McCombie, S. W.; Baroudy, B. M.; Smith, S. O.; Dragic, T.; Moore, J. P.; Sakmar, T. P.; Interaction of small molecule inhibitors of HIV-1 entry with CCR5. *Virology* **2006**, *349*, 41−54.
(51) Zhou, N.; Luo, Z.; Hall, J. W.; Luo, J.; Han, X.; Huang, Z.; Molecular modeling and site-directed mutagenesis of CCR5 reveal residues critical for chemokine binding and signal transduction. *Eur. J. Immunol.* **2000**, *30*, 164−173.
(52) Dragic, T.; Trkola, A.; Thompson, D. A.; Cormier, E. G.; Kajumo, F. A.; Maxwell, E.; Lin, S. W.; Ying, W.; Smith, S. O.; Sakmar, T. P.; Moore, J. P. A binding pocket for a small molecule inhibitor of HIV-1 entry within the transmembrane helices of CCR5. *Proc. Natl. Acad. Sci., U.S.A.* **2000**, *97*, 5639−5644.
(53) Baba, M.; Nishimura, O.; Kanzaki, N.; Okamoto, M.; Sawada, H.; Iizawa, Y.; et al. A small-molecule, nonpeptide CCR5 antagonist with highly potent and selective anti-HIV-1 activity. *Proc. Natl. Acad. Sci., U.S.A.* **1999**, *96*, 5698−5703.
(54) Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 0022600, 2002.
(55) Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 0022599, 2002.
(56) Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 00234745, 2002.
(57) Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 055876, 2003.
(58) Bridger, G.; Skerlj, R.; Kaller, A.; Harwig, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y.; Schools, D.; Smith, C. D.; Di Fluir, R. M. U.S. Patent 7091217, 2004.
(59) Ichiyama, K.; Yokohama-Kumakura, S.; Tanaka, Y.; Tanaka, R.; Hirose, K.; Bannai, K.; Edamatsu, T.; Yanaka, M.; Niitani, Y.; Miyako-Kurosaki, N.; Takaku, H.; Koyanagi, Y.; Yamamoto, N. A duodenally absorbable CXC chemokine receptor 4 antagonist, KRH-1636, exhibits a potent and selective anti-HIV-1 activity. *Proc. Natl. Acad. Sci*., *U.S.A*. **2003**, *100*, 4185−4190.
(60) Murakami, T.; Yoshida, A.; Tanaka, R.; Mitsuhashi, S.; Hirose, K.; Yanaka, M.; Yamamato, N.; Tanaka, Y. KRH-2731: An Orally Bioavailable CXCR4 Antagonist Is a Potent Inhibitor of HIV-1 Infection. In *2004 Antivirals Pipeline Report*; Camp, R., Ed.; Proceedings of the 11th Conference on Retroviruses and Opportunistic

Infection, San Francisco CA, Feb. 8−11, 2004; Treatment Action Group: San Francisco, CA, 2004; Abstract No. 541.
(61) Yamazaki, T.; Saitou, A.; Ono, M.; Yokohama, S.; Bannai, K.; Hiroswe, K.; Yanaka, M. World Patent WO 029218, 2003.
(62) Yamazaki, T.; Kikumoto, S.; Ono, M.; Saitou, A.; Takahashi, H.; Kumakura, S.; Hirose, K. World Patent WO 024697, 2004.
(63) Bridger, G. J.; Skerlj, R. T.; Padmanabhan, S.; Martellucci, S. A.; Henson, G. W.; Struyf, S.; Witvrouw, M.; Schols, D.; De Clercq, E. Synthesis and Structure−Activity Relationships of Phenylenebis-(methylene)-Linked Bis-azamacrocycles That Inhibit HIV-1 and HIV-2 Replication by Antagonism of the Chemokine Receptor CXCR4. *J. Med. Chem.* **1999**, *42*, 3971−3981.
(64) De Clercq, E. Inhibition of HIV Infection by Bicyclams, Highly Potent and Specific CXCR4 Antagonists. *Mol. Pharmacol.* **2000**, *57*, 833−839.
(65) Esté, J. A.; Cabrera, C.; De Clercq, E.; Struyf, S.; Van Damme, J.; Bridger, G.; Skerlj, R. T.; Abrams, M. J.; Henson, G.; Gutierrez, A.; Clotet, B.; Schols, D. Activity of Different Bicyclam Derivatives against Human Immunodeficiency Virus Depends on Their Interaction with the CXCR4 Chemokine Receptor. *Mol. Pharmacol.* **1999**, *55*, 67−73.
(66) Egberink, H. F.; De Clercq, E.; Van Vliet, A. L. W.; Balzarini, J.; Bridger, G. J.; Henson, G.; Horzinek, M. C.; Schols, D. Bicyclams, selective antagonists of the human chemokine receptor CXCR4, potently inhibit feline immunodeficiency virus replication. *J. Virol.* **1999**, *73*, 6346−6352.
(67) Hatse, S.; Princen, K.; De Clercq, E.; Rosenkilde, M. M.; Schwartzb, T. W.; Hernandez-Abad, P. E.; Skerlj, R. T.; Bridger, G. J.; Schols, D. AMD3465, a monomacrocyclic CXCR4 antagonist and potent HIV entry inhibitor. *Biochem. Pharmacol.* **2005**, *70*, 752−761.
(68) Princen, K.; Hatse, S.; Vermeire, K.; Aquaro, S.; De Clercq, E.; Gerlach, L.-O.; Rosenkilde, M.; Schwartz, T. W.; Skerlj, R.; Bridger, G.; Schols, D. Inhibition of Human Immunodeficiency Virus Replication by a Dual CCR5/CXCR4 Antagonist. *J. Virol.* **2004**, *78*, 12996−13006.
(69) Tamamura, H.; Araki, T.; Ueda, S.; Wang, Z.; Oishi, S.; Esaka, A.; Trent, J. O.; Nakashima, H.; Yamamoto, N.; Peiper, S. C.; Otaka, A.; Fujii, N. Identification of novel low molecular weight CXCR4 antagonists by structural tuning of cyclic tetrapeptide scaffolds. *J. Med. Chem.* **2005**, *48*, 3280−3289.
(70) Rosenkilde, M. M.; Gerlach, L. O.; Hatse, S.; Skerlj, R. L.; Schols, D.; Bridger, G.; Schwartz, T. W. Molecular mechanism of action of monociclam versus biciclam non-peptide antagonist in the CXCR4 chemokine receptor. *J. Biol. Chem.* **2007**, *282*, 27354−27365.
(71) Palani, A.; Shapiro, S.; Clades, J. W.; Greenlee, W. J.; Blythin, D.; Cox, K.; Wagner, N. E.; Strizki, J.; Baroudy, B. M.; Dan, N. Biological evaluation and interconversion studies of rotamers of SCH 351125, an orally bioavailable CCR5 antagonist. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 705−708.
(72) Tsamis, F.; Gavrilov, S.; Kajumo, F.; Seibert, C.; Kuhmann, S.; Ketas, T.; Trkola, A.; Palani, A.; Clader, J. W.; Tagat, J. R.; McCombie, S.; Baroudy, B.; Moore, J. P.; Sakmar, T. P.; Dragic, T. Analysis of the Mechanism by Which the Small-Molecule CCR5 Antagonists SCH-351125 and SCH-350581 Inhibit Human Immunodeficiency Virus Type 1 Entry. *J. Virol.* **2003**, *77*, 5201−5208.
(73) Billick, E.; Seibert, C.; Pugach, P.; Ketas, T.; Trkola, A.; Endres, M. J.; Murgolo, N. J.; Coates, E.; Reyes, G. R.; Baroudy, B. M.; Sakmar, T. P.; Moore, J. P.; Kuhmann, S. E. The Differential Sensitivity of Human and Rhesus Macaque CCR5 to Small-Molecule Inhibitors of Human Immunodeficiency Virus Type 1 Entry Is Explained by a Single Amino Acid Difference and Suggests a Mechanism of Action for These Inhibitors. *J. Virol.* **2004**, *78*, 4134−4144.
(74) Maeda, K.; Yoshimura, K.; Shibayama, S.; Habashita, H.; Tada, H.; Sagawa, K.; Mikayawa, T.; Auki, M.; Fukushima, D.; Mitsuya, H. Novel Low Molecular Weight Spirodiketopiperazine Derivatives Potently Inhibit R5 HIV-1 Infection through Their Antagonistic Effects on CCR5. *J. Biol. Chem.* **2001**, *276*, 35194−35200.
(75) Shibayama, S.; Sagawa, K.; Watanabe, N.; Takeda, K.; Tada, H.; Fukushima, D. World Patent WO 2004054616, 2004.
(76) Takaoka, Y.; Okamoto, M.; Genba, Y. World Patent WO 2004026874, 2004.
(77) Takaoka, Y.; Nishizawa, R.; Shibayama, S.; Sagawa, K.; Matsuo, M. Y. World Patent WO 2002074770, 2002.
(78) Imawaka, H.; Shibayama, S.; Takaoka, Y. World Patent WO 2003035074, 2003.
(79) Cumming, J.; Tucker, H. World Patent WO 2003042177, 2003.
(80) Cumming, J. World Patent WO 2003042178, 2003.
(81) Cumming, J. World Patent WO 2003080574, 2003.
(82) Cumming, J.; Winter, J. World Patent WO 2004018425, 2004.
(83) Burrows, J.; Cumming, J. World Patent WO 2002076, 2002.
(84) Willouhby, C. W.; Rosauer, K. G.; Hale, J. J.; Budhu, R. J.; Mills, S. G.; Chapman, K. T.; MacCoss, M.; Malkowitz, L.; Springer, M. S.; Gould, S. L.; DeMartino, J. A.; Siciliano, S. J.; Cascieri, M. A.; Carella, A.; Catver, G.; Colmes, K.; Schlief, W. A.; Danzeisen, R.; Hazuda, D.; Kessler, J.; Lineberger, J.; Miller, M.; Emini, E. A. 1,3,4 Trisubstituted pyrrolidine CCR5 receptor antagonists bearing 4-aminoheterocycle substituted piperidine side chains. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 427−431.
(85) Kazmierski, W. M.; Aquino, C. J.; Bifulco, N.; Boros, E. E.; Chauder, B. A.; Chong, P. Y.; Duan, M.; Deanada, F., Jr.; Koble, C. S.; Malean, E. W.; Peckham, J. P.; Perkins, A. C.; Thompson, J. B.; Vanderwall, D. World Patent WO 2004054974, 2004.
(86) Duan, M., Kazmierski, W. M.; Aquino, C. J. World Patent WO 200405481, 2004.
(87) Peckham, J. P.; Aquino, C. J.; Kazmierski, W. M. World Patent WO 2004055010, 2004.
(88) Aquino, C. J.; Chong, P. Y.; Duan, M.; Kazmierski, W. M. World Patent WO 2004055011, 2004.
(89) Youngman, M.; Kazmierski, W. M.; Yang, H.; Aquino, C. J. World Patent WO 2004055012, 2004.
(90) Yang, H.; Kazmierski, W. M.; Aquino, C. J. World Patent WO 2004055016, 2004.
(91) Aramaki, Y.; Seto, M.; Okawa, T.; Oda, T.; Kanzaki, N.; Shiraishi, M. Synthesis of 1-benzothiepine and 1-benzazepine derivatives as orally active CCR5 antagonists. *Chem. Pharm. Bull*. **2004**, *52*, 254−258.
(92) Seto, M.; Aramaki, Y.; Okawa.; T.; Miyamoto, N.; Aikawa, K.; Kanzaki, N.; Shiraishi, M. Orally active antagonists as Anti-HIV-1 Agents: Synthesis and Biological Activity of 1-Benzothiepine 1,1-Dioxide and 1-Benzazepine derivatives containing a tertiary Amine Moiety. *Chem. Pharm. Bull.* **2004**, *52*, 577−590.
(93) Perros, M.; Price, D. A.; Stammen, B. L. C.; Wood, A. World Patent WO 2003084954, 2003.
(94) Basford, P. A.; Stephenson, P. T.; Taylor, S. C. J.; Wood, A. World Patent WO 2003084954, 2003.
(95) Armour, D. R.; Price, D. A.; Stammen, B. L. C.; Wood, A.; Perros, M.; Edwards, M. P. World Patent WO 2000038680, 2000.
(96) Rusconi, S.; Scozzafava, A.; Mastrolorenzo, A.; Supuran, T. C. New Advances in HIV Entry Inhibitors Development. *Curr. Drug Targets Infect. Disord.* **2004**, *4*, 339−355.
(97) Imamura, S.; Kurasawa, O.; Nara, Y.; Ichikawa, T.; Nishikawa, Y.; Iida, T.; Hashiguchi, S.; Kanzaki, N.; Lizawa, Y.; Baba, M.; Sugihara, Y. CCR5 antagonists as anti-HIV-1 agents. Part 2: Synthesis and biological evaluation of *N*-[3-(4-benzylpiperidin-1-yl)propyl]-*N*,*N*′-diphenylureas. *Bioorg. Med. Chem.* **2004**, *12*, 2295−2306.
(98) Imamura, S.; Ishihara, Y.; Hattori, T.; Kurasawa, O.; Matsushita, Y.; Sugihara, Y.; Kanzaki, N.; Lizawa, Y.; Baba, M.; Hashiguchi, S. CCR5 antagonists as Anti-HIV-1 agents. 1. Synthesis and Biological Evaluation of 5-oxopyrrolidine-3-carboxamide Derivatives. *Chem. Pharm. Bull.* **2004**, *52*, 63−73.
(99) Debnath, A. K. Generation of Predictive Pharmacophore Models for CCR5 Antagonists: Study with Piperidine- and Piperazine-Based Compounds as a New Class of HIV-1 Entry Inhibitors. *J. Med. Chem.* **2003**, *46*, 4501−4515.
(100) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.
(101) *Maybride Bringing life to drug discovery*, Maybridge Databases Autumn 2005; Fisher Scientific International: England, 2005.
(102) Labute, P. Flexible Alignment of Small Molecules. Chemical Computing Group, Inc.: Montreal, Canada, 2004. (Available via the Internet at http://www.chemcomp.com/journal/malign.htm, accessed Dec. 23, 2007.)
(103) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68*, 76−90.
(104) Ritchie, D. W.; Kemp, G. J. L. Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces. *J. Comput. Chem.* **1999**, *20*, 383−395.
(105) Ritchie, D. W.; Kemp, G. J. L.; Protein docking using spherical polar Fourier correlations. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 178−194.
(106) Wang, R.; Wang, S. How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422−1426.
(107) Lin, J.; Clark, T. An analytical, variable resolution, complete description of static molecules and their intermolecular binding properties. *J. Chem. Inf. Model.* **2005**, *45*, 1010−1016.
(108) Grant, A. J.; Pickup, B. T. A fast method of molecular shape comparison: a simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1996**, *17*, 1653−1659.
(109) *OMEGA*, version 2.1.0; OpenEye Scientific Software, Inc.: Santa Fe, NM., 2006.
(110) Warren, G. L.; Andrews, C. V.; Capelli, A.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head,

M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912−5931.

(111) Liang, X.; Parkinson, J. A.; Weishäupl, M.; Gould, R. O.; Paisey, S. J.; Park, H.; Hunter, T. M.; Blindauer, C. A.; Parsons, S.; Sadler, P. J. Structure and Dynamics of Metallomacrocycles: Recognition of Zinc Xylyl-Bicyclam by an HIV Coreceptor. *J. Am. Chem. Soc.* **2002**, *124*, 9105−9112.

(112) Trent, J. O.; Wang, Z. X.; Murray, J. L.; Shao, W.; Tamamura, H.; Fujii, N.; Peiper, S. C. Lipid Bilayer Simulations of CXCR4 with Inverse Agonist and Weak Partial Agonist. *J. Biol. Chem.* **2003**, *278*, 47136−47144.

(113) McDonald, I. K.; Thornton, J. M. Satisfying Hydrogen Bonding Potential in Proteins. *J. Mol. Biol.* **1994**, *238*, 777−793.

(114) Dragic, T.; Trkola, A.; Lin, S. W.; Nagashima, K. A.; Kajumo, F.; Zhao, L.; Olson, W. C.; Wu, L.; Mackay, C. R.; Allaway, G. P.; Sakmar, T. P.; Moore, J. P.; Maddon, P. J. Amino-terminal substitutions in the CCR5 co-receptor impair gp120 binding and human immunodeficiency virus type 1 entry. *J. Virol.* **1998**, *72*, 279−285.

(115) Rucker, J.; Samson, M.; Doranz, B. J.; Libert, F.; Berson, J. F.; Yi, Y.; Smyth, R. J.; Collman, R. G.; Broder, C. C.; Vassart, G.; Doms, R. W.; Parmentier, M. Regions in beta-chemokine receptors CCR5 and CCR2b that determine HIV-1 cofactor specificity. *Cell* **1996**, *87*, 437−446.

(116) Fano, A.; Ritchie, D. W.; Carrieri, A. Modelling the structural basis of human CCR5 chemokine receptor function: from homology model-building and molecular dynamics validation to agonist and antagonist docking. *J. Chem. Inf. Model.* **2006**, *46*, 1223−1235.

(117) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of Docking Performance: Comparative Data on Docking Algorithms. *J. Med. Chem.* **2004**, *47*, 558−565.

(118) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.

(119) Gutiérrez-de-Terán, H.; Pastor, M.; Centeno, N. B.; Aqvist, J.; Sanz, F. Comparative analysis of putative agonist binding modes in the human A1 adenosine receptor. *ChemBioChem* **2004**, *5*, 841−849.

(120) Shields, G. C.; Laughton, C. A.; Orozco, M. Molecular dynamics simulations of the d(T·A·T) triple helix. *J. Am. Chem. Soc.* **1997**, *119*, 7463−7469.

(121) Evers, A.; Hessler, G.; Matter, H.; Klabunde, T. Virtual Screening of Biogenic Amine-Binding G-Protein Coupled Receptors: Comparative Evaluation of Protein- and Ligand-Based Virtual Screening Protocols. *J. Med. Chem.* **2005**, *48*, 5448−5465.

(122) Rabal, O.; Schneider, G.; Borrell, J. I.; Teixidó, J. Structure-Based Virtual Screening of FGFR Inhibitors. Cross-Decoys and Induced-Fit Effect. *BioDrugs* **2007**, *21*, 31−45.

(123) Teixidó, J.; Borrell, J. I.; Nonell, S.; Batllori, X.; Pettersson, S.; Ros, L.; Puig de la Bellacasa, R.; Rabal, M. O.; Pérez-Nueno, V.; Esté, J.; Clotet, I.; Armand-Ugón, M.; Nuevos sistemas polinitrogenados como agentes anti-VIH. Spanish Patent ES200602764, Oct. 26, 2006.

(124) Pettersson, S.; Clotet-Codina, I.; Esté, J. A.; Borrell, J. I.; Teixidó, J. Recent advances in combinatorial chemistry applied to development of anti-HIV drugs. *Mini Rev. Med. Chem.* **2006**, *6*, 91−108.

(125) Hawkins, P. D. C.; Skillman, G. A.; Nicholls, A. Comparison of Shape Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74−82.

(126) Castonguay, L. A.; Weng, Y.; Adolfsen, W.; Di Salvo, J.; Kilburn, R.; Caldwell, C. G.; Daugherty, B. L.; Finke, P. E.; Hale, J. J.; Lynch, C. L.; Mills, S. G.; MacCoss, M.; Springer, M. S.; DeMartino, J. A. Binding of 2-Aryl-4-(piperidin-1-yl)butanamines and 1,3,4-Trisubstituted Pyrrolidines to Human CCR5: A Molecular Modeling Guided Mutagenesis Study of the Binding Pocket. *Biochemistry* **2003**, *42*, 1544−1550.

CI700415G