





## Capítulo 5.

### Cribado por *docking*

Siguiendo con el esquema de filtros sucesivos aplicados en una cascada de cribado virtual, tras la aproximación *ligand-based* establecida en el capítulo 4, se decide estudiar y validar una aproximación *structure-based* a partir de métodos de *docking* que permita evaluar los compuestos incluidos en las quimiotecas BIB\_Oxo, BIB\_Amino y BIB\_Hidro descritas en el capítulo 3.

El gran número de estructuras disponibles de proteína tirosina quinasa (PTKs) con una buena resolución ha demostrado que esta familia de proteínas constituye una familia adecuada para el cribado virtual basado en la estructura.<sup>379-380</sup> También el uso de modelos por homología se ha mostrado una aproximación útil en esta familia.<sup>68,381-382</sup>

En este caso se han escogido como dianas terapéuticas los receptores tirosina quinasa (RTKs) PDGFR, EGFR, y FGFR, para las que los sistemas pirido[2,3-*d*]pirimidínicos muestran actividad inhibitoria con mayor o menor selectividad. Para los dos últimas existen estructuras depositadas en el *Protein Data Bank*, tanto en su forma apo como cristalizadas con diferentes inhibidores. Sin embargo, no existe actualmente (junio 2006) ninguna estructura del dominio tirosina quinasa de PDGFR depositado en dicha base de datos, por lo que se aborda también la construcción de un modelo por homología de dicha estructura.

Previamente al cribado virtual, se ha determinado la capacidad de reproducir los modos de unión para una serie de ligandos co-cristalizados en las dos proteínas con estructuras resueltas por rayos-X. El motivo es que normalmente se acepta que un cribado virtual basado en estructuras incorrectamente “*dockeadas*” (acopladas) debería alcanzar factores de enriquecimiento pobres, pese a que algunos autores han encontrado que los resultados de un cribado virtual con *docking* no siempre están relacionados con la calidad de los modos de unión predichos.<sup>383</sup> Además del modo de unión, se ha explorado la capacidad del *docking* como herramienta para predecir el sitio de unión (*docking* ciego o *blind docking*), una aplicación para la que los programas de *docking* no han sido especialmente optimizados.

Cuando se establece un protocolo apropiado de cribado virtual basado *docking* para una diana biológica particular, el primer paso conlleva la identificación de la mejor combinación de un algoritmo de *docking* y una función de *scoring*. En este sentido, se han publicado numerosos estudios que evalúan diversos aspectos de los programas de *docking* como comparaciones de las funciones de *scoring*<sup>384-387</sup>, evaluaciones de los algoritmos de *docking*<sup>388</sup>, análisis de composiciones de los dos factores<sup>389-390</sup> y la inclusión de *consensus scoring*.<sup>215,383,391-392</sup>

No existe actualmente una combinación óptima aplicable a todas las dianas terapéuticas, mostrando unas u otras combinaciones diferentes habilidades en el *docking*. De hecho, en ausencia de un ajuste especial de una función de *scoring* para una diana biológica, resulta imposible *a priori* determinar si un programa de *docking* será exitoso o no en un cribado virtual frente a una diana biológica en particular. En este trabajo, se han utilizado dos programas basados en funciones de *scoring* empíricas: GOLD<sup>80</sup> y AUTODOCK<sup>81</sup> (apartado 1.4.1), para los que diversos parámetros iniciales han sido analizados.

Además de seleccionar la mejor combinación de algoritmo de *docking* y función de *scoring*, otro de los puntos críticos en un cribado virtual es la elección de la estructura de la proteína a usar cuando varias están disponibles. Como se ha introducido, la mayor parte de programas de *docking* asumen la rigidez del sitio activo, debido al coste computacional que supone la inclusión de la flexibilidad de la proteína.

El impacto de la estructura de la proteína en el funcionamiento del *docking* ha sido estudiado principalmente en el contexto de la predicción de los modos de unión de ligandos.<sup>385,387,393-395</sup> Así, para un conjunto de estructuras de una proteína particular, se intercambian los diferentes ligandos co-cristalizados que presentan, generándose los denominados *cross-decoys* y se evalúa su capacidad de reproducir los modos de unión en estructuras diferentes de su estructura “nativa”. Recientemente, se ha publicado un estudio que examina este aspecto para la quinasa CDK2<sup>396</sup>, constatándose la elevada influencia que tiene la proteína en el éxito o no del *docking*. Esto es particularmente problemático para la familia de las quinasas, debido a la elevada flexibilidad de éstas generada por efectos de inducción (*induced-fit effects*) tras la unión del ligando.<sup>330</sup>

Más allá de la predicción del modo de unión, la influencia de la estructura de la proteína en aplicaciones de cribado virtual ha sido mucho menos analizada. Mc.Govern y Shoichet<sup>397</sup> compararon las conformaciones holo, apo y procedentes de modelos por homología para una serie de sitios de unión, mostrando la superioridad general en recuperar ligandos activos de la forma holo frente a la apo. Particularmente, Cavasotto *et al*<sup>379</sup> analizan este aspecto en profundidad en cuatro quinasas (cAPK, CDK2, P38 y LCK).

Por ello, a la hora de establecer un protocolo de cribado virtual en este trabajo, se ha analizado también la influencia de la estructura de la proteína escogida, tanto en la reproducción de los modos de unión como en el funcionamiento del cribado virtual. El objetivo, establecer si existe una conformación “preferida” para ser usada en el cribado virtual o si se puede alcanzar un *consensus* a partir de varias estructuras. Recientemente se han desarrollado métodos computacionales que permiten recoger la flexibilidad del receptor de una manera eficiente, basados en generar múltiples conformaciones del mismo.<sup>398</sup> Estos métodos no han sido tratados en este trabajo.

Según lo expuesto, para las dianas consideradas, se analizan los siguientes puntos:

- i) Predicción del modo de unión para los ligandos nativos de una estructura de la proteína en particular y para los ligandos co-cristalizados en otras estructuras (*cross-decoys*) inicialmente posicionados en el *binding pocket*. Realizado con los programas AUTODOCK y GOLD.
- ii) Predicción del sitio de unión y modo de unión para una serie de ligandos conocidos situados inicialmente en la superficie de la proteína. (*Blind docking* o *docking* ciego). Se utiliza el programa AUTODOCK. GOLD no se usa en este caso, ya que requiere que el usuario indique el tamaño aproximado y la localización del sitio de unión del ligando.<sup>80</sup>
- iii) Cribado virtual retrospectivo de una base de datos de compuestos activos. Establecimiento del protocolo óptimo y validación del mismo. Realizado con los programas AUTODOCK y GOLD.

Los resultados del cribado virtual se comparan con los factores de enriquecimiento obtenidos con las aproximaciones *ligand-based* descritas en el capítulo 4.

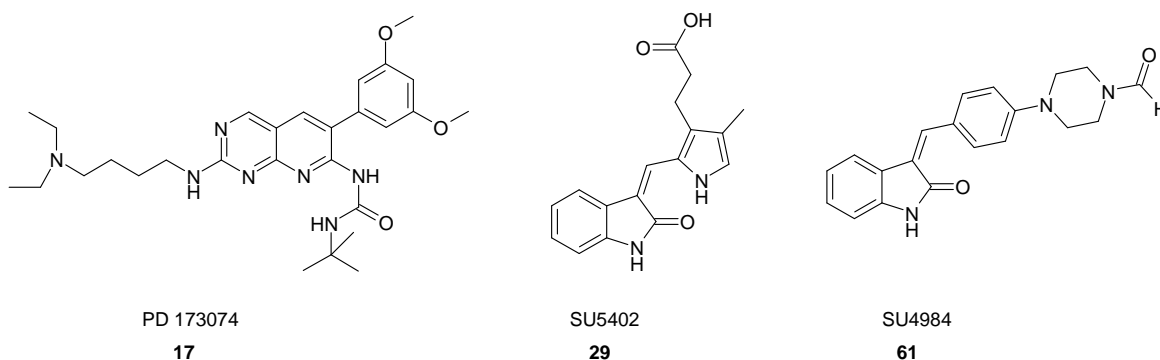
El capítulo finaliza con la aplicación prospectiva del protocolo de *docking* escogido para las quimiotecas BIB\_Oxo, BIB\_Amino y BIB\_Hidro.

## 5.1. *Docking* frente a FGFR

En este apartado se analizan los citados tres puntos (predicción del modo de unión, predicción de sitio de unión y cribado virtual) frente a FGFR.

### 5.1.1. Preparación de las estructuras cristalinas de FGFR

Como se ha descrito en el apartado 4.2, existen tres estructuras cristalizadas del dominio quinasa de FGFR1 complejadas con tres inhibidores: PD173074 (entrada PDB 2fgi, resolución 2.5 Å), SU5402 (entrada PDB 1fgi, resolución 2.5 Å) y SU4984 (entrada PDB 1agw, resolución 2.4 Å). PD173074 presenta actividad frente a FGFR en el rango nanomolar, mientras que las dos restantes son dos órdenes de magnitud menos activas (Figura 5.1).



**Figura 5.1.** Estructuras de inhibidores co-cristalizados con FGFR.

Para cada complejo, se eliminan las aguas y se añaden los átomos de hidrógeno. En cada caso, se trabaja únicamente con una de las dos cadenas del dímero resueltas, la cadena que incluye más residuos determinados estructuralmente. La posición de los átomos de hidrógeno se relaja usando el *force field* AMBER incluido en MOE<sup>38</sup>, seguido de una minimización del ligando con el *force field* MMFF94 usando tres métodos consecutivos (*Steepest Descent*, *Conjugate Gradient* y *Truncated Newton*) para alcanzar una optimización energética con un gradiente de RMS inferior a 0.05.

En el caso del *docking* ciego, se genera un conjunto de posiciones iniciales (*decoys*) alejadas del sitio de unión, es decir, los ligandos se colocan aleatoriamente en la superficie de la proteína y se minimizan bajo las mismas condiciones descritas.

Para generar los seis *cross-decoys* (estructuras proteína-inhibidor intercambiadas), se superponen las tres estructuras del receptor y las posiciones de los otros dos ligandos no nativos se optimizan dentro del *binding pocket* “extraño”, siguiendo el procedimiento descrito. De nuevo, se generan posiciones para estas estructuras cruzadas en la superficie de la proteína (*docking* ciego).

Los grupos ionizables (aminas, ácidos carboxílicos, fosfatos, amidinas) se asumen ionizados a pH fisiológico tanto para la proteína como para los ligandos, excepto si se consideran datos experimentales. Particularmente, el grupo butilamino de PD173074 se mantiene cargado y para SU5402, se mantiene protonado el grupo ácido carboxílico.

### 5.1.2. Procedimiento para la predicción del modo de unión: estructuras nativas y *cross-decoys*

En primer lugar, se describe el procedimiento aplicado en cada uno de los programas utilizados (AUTODOCK y GOLD). Estos parámetros, se han aplicado posteriormente para EGFR y PDGFR, excepto algunas modificaciones realizadas, debidamente especificadas.

- **AUTODOCK.** Se ha utilizado tanto para la predicción del modo de unión como para el *docking* ciego. En ambos casos, las dimensiones de la malla tridimensional son 181×181×181, situándose centradas en el sitio de unión del ligando y con un espaciado entre los puntos de 0.375 Å. Este tamaño de la malla se escoge inicialmente lo suficientemente grande como para acomodar a la estructura completa de la proteína, permitiendo acceder al ligando a toda la superficie. De este modo, el espacio de búsqueda no queda exclusivamente reducido al sitio de unión conocido. Como se ha comentado en el apartado 1.4.1.1, AUTODOCK calcula, mediante el programa AUTOGRID, potenciales de afinidad atómica para cada uno de los tipos atómicos definidos para el sustrato. En este trabajo, se han analizado dos *settings* iniciales para tratar los tipos atómicos, aparte del mapa de potencial electrostático:

- ◆ El set típico, que utiliza los siguientes tipos atómicos: C (carbono alifático), A (carbono aromático), N, O y H.
- ◆ Un segundo set que además de incluir a los anteriores, define también los átomos de Br, Cl, P y S. La importancia de incluir parámetros de puentes de hidrógeno especiales en el *docking* frente a tirosina quinasas ha sido previamente reportada.<sup>216,399</sup> Debido a la interacción clave N-H...acceptor, conservada en muchos inhibidores ATP-competitivos, se opta por parametrizar la posibilidad de dicha interacción para el átomo de nitrógeno. Para ello, es necesario distinguir en la asignación de los tipos atómicos entre los nitrógenos aceptores (N) y dadores (n). Así, se implementa un mapa de potencial adicional en los cálculos de AUTOGRID, cuyos parámetros se extraen de la *web* de AUTODOCK.<sup>400</sup>

Para la proteína, únicamente se añaden los hidrógenos polares (modelo *united-atom*) y se asignan las cargas Kollman, tal y como se encuentran incluidas en MOE. Para los inhibidores, se mantienen todos los hidrógenos. La influencia de las cargas parciales de los ligandos se evalúa en conjunción con los tipos atómicos mencionados, rindiendo dos *settings* iniciales:

- ◆ El set de tipos atómicos tradicionales (C,A,N,O,H) y las cargas Gasteiger-Marsili<sup>171</sup>, implementadas en MOE.
- ◆ El set de tipos atómicos adicionales (C,A,N,n,O,H,Br,Cl,F,S,P) junto a las cargas semiempíricas AM1, calculadas con el paquete MOPAC.<sup>266</sup>

En cada *docking*, se calculan 100 conformeros independientes, correspondiendo cada uno de ellos al resultado de una optimización iniciada con diferentes valores aleatorios. En el análisis de los resultados, se toman como referencia las coordenadas cristalográficas del ligando cristalizado, agrupándose las 100 conformaciones obtenidas en *clusters* según el valor de RMS de superposición entre ellas, con una tolerancia de 1.0 Å. Cada *docking* se repite 5 veces. En el proceso de búsqueda conformacional, se aplica el algoritmo genético lamarckiano (LGA) y el método pseudo-Solis y Wets para la minimización posterior de la conformación, con los parámetros por defecto del programa.<sup>81</sup> El parámetro *Intelec* se activa, de manera que la energía electrostática interna del ligando afecta a la energía de

*docking* (*docked energy*) que guía el proceso de optimización, aunque no a la energía de unión del ligando (*binding free energy*).

- **GOLD.** En este punto del trabajo, se ha utilizado la versión 2.2 de GOLD. GOLD requiere que el usuario indique el tamaño aproximado y la localización del sitio de unión del ligando. Por lo tanto, únicamente es posible realizar un *docking* con conocimiento del sitio de unión y no un *docking* ciego. Así, se define el sitio activo compuesto de los átomos de la proteína dentro de 20 Å a partir de los más próximos al centro del *binding site*. Se activa el algoritmo de detección de cavidad (*cavity-detection algorithm*), de manera que el cálculo se confina a las regiones cóncavas en la vecindad del sitio de unión. En pruebas previas realizadas, si se desactiva la opción de este algoritmo de detección, se asigna un número muy inferior de átomos como sitio activo, influyendo negativamente el *docking*. Durante el *docking*, se permite una cierta flexibilidad en los átomos de hidrógeno (de –OH y –NH<sub>3</sub>) de las cadenas laterales de Ser, Thr y Lys.

GOLD utiliza también un algoritmo genético (GA) para explorar los modos de unión posibles. Como en AUTODOCK, se requieren explícitamente 100 conformaciones independientes por experimento de *docking*, consistiendo cada ejecución de un máximo de 100.000 operaciones. Esta parametrización corresponde al *setting* exhaustivo de la referencia [80], excepto por el hecho de que se permite una terminación temprana de la optimización cuando las tres mejores soluciones del *docking* se encuentran en un margen de RMSD de 1.5 Å de superposición entre ellas. El resto de parámetros se mantiene por defecto: un tamaño de población de 100, una presión selectiva de 1.1, 5 islas (*island model*) y un tamaño de *niche* de 2 (véase apartado 1.9.2). Se establecen unos valores de *cutoff* de 2.5 Å para los puentes de hidrógeno y de 4.0 Å para los contactos de vdW. Como funciones de *scoring*, se utilizan separadamente las funciones GOLDScore y CHEMScore (apartados 1.4.1.2 y 1.4.1.3, respectivamente). De nuevo, cada *docking* se repite 5 veces.

Contrariamente a AUTODOCK, GOLD utiliza un modelo de todos los átomos (*all-atom model*), de manera que en este caso se mantienen todos los hidrógenos (polares y no polares) en el fichero de la proteína. Además, GOLD ignora tanto las cargas formales como las cargas parciales asignadas.

Para el cálculo de la RMSD de superposición de la mejor conformación con la estructura cristalográfica, únicamente se consideran los átomos pesados (*heavy atoms*), utilizándose el programa *smart\_rms* del paquete de *software* GOLD.

### 5.1.3. Resultados de la predicción del modo de unión en FGFR: estructuras nativas y *cross-decoys*

En la Tabla 5.1 se muestran los resultados obtenidos para el *docking* y el *cross-docking* de las posiciones (*decoys*) inicialmente situadas en el sitio de unión para los cuatro protocolos descritos en el apartado anterior:

- i) AUTODOCK con tipos atómicos estándar y las cargas Gasteiger-Marsili
- ii) AUTODOCK con los mapas de afinidad adicionales y cargas AM1.
- iii) GOLD con GOLDScore y
- iv) GOLD con CHEMScore

Estos resultados se expresan como la raíz de la desviación cuadrática media (RMSD, Å) entre la conformación cristalina y la mejor de las conformaciones obtenidas (*poses*) en cada ejecución. Aquellas conformaciones calculadas con desviaciones  $\leq 2$  Å se consideran correctas.<sup>385,389</sup> En la Tabla 5.1, los resultados dentro de 2 Å aparecen sombreados en gris oscuro, aquellos

comprendidos entre 2 y 4 Å están sombreados en gris claro y aquellos con una RMSD superior a 4 Å tienen un fondo blanco. Los resultados en negrita corresponden a complejos nativos, es decir, al par ligando-proteína co-cristalizado. Los valores se presentan como el promedio de cinco repeticiones, con su desviación estándar en paréntesis. La referencia a la estructura “2fgi-transfer” se describe más adelante.

**Tabla 5.1.** Resultados de RMSD (Å) y sus desviaciones estándar (paréntesis) para las posiciones inicialmente situadas en el sitio de unión en el *docking* y el *cross-docking* frente a las tres estructuras resueltas de FGFR para los cuatro protocolos.

		PD173074	SU5402	SU4984
AUTODOCK Tipos atómicos estándar Cargas Gasteiger-Marsili	2fgi	<b>1.5 (0.2)</b>	14.3 (2.8)	17.5 (0.6)
	2fgi-transfer	1.8 (0.3)	12.6 (6.0)	16.5 (1.6)
	1fgi	9.7 (0.6)	<b>2.3 (0.2)</b>	4.7 (0.3)
	1agw	9.7 (4.8)	15.5 (0.3)	<b>11.9 (6.1)</b>
AUTODOCK Mapas de afinidad adicionales Cargas AM1	2fgi	<b>1.7 (0.3)</b>	15.5 (0.5)	17.7 (0.2)
	2fgi-transfer	1.6 (0.2)	15.9 (1.1)	14.5 (4.7)
	1fgi	7.4 (2.0)	<b>2.5 (0.2)</b>	5.1 (0.9)
	1agw	10.8 (0.3)	15.5 (0.1)	<b>9.9 (5.0)</b>
GOLD GOLDScore	2fgi	<b>2.2 (0.4)</b>	7.6 (0.2)	5.0 (0.1)
	2fgi-transfer	1.9 (0.1)	5.2 (0.8)	4.6 (0.1)
	1fgi	10.7 (0.2)	<b>2.7 (0.2)</b>	2.8 (0.3)
	1agw	10.8 (0.5)	6.3 (0.2)	<b>6.8 (0.1)</b>
GOLD CHEMscore	2fgi	<b>2.4 (0.5)</b>	1.1 (0.1)	4.2 (0.2)
	2fgi-transfer	5.7 (0.3)	1.0 (0.1)	4.6 (0.1)
	1fgi	11.5 (0.8)	<b>1.3 (0.1)</b>	1.7 (0.7)
	1agw	12.0 (4.3)	1.2 (0.1)	<b>1.9 (0.0)</b>

En primer lugar se discuten los resultados de la predicción del modo de unión para los ligandos nativos inicialmente posicionados en el sitio de unión (*negrita* en Tabla 5.1). Para los tres complejos, al menos un programa es capaz de reconocer la orientación y el modo de unión dentro de 2 Å. Tras inspeccionar visualmente las conformaciones resultantes dentro de 2.5 Å para PD173074 y SU5402, se aprecia que las interacciones más importantes se conservan. Los resultados obtenidos con GOLDScore para los complejos 1fgi y 2fgi están de acuerdo con los reportados por Kontoyianni y colaboradores<sup>388</sup>, quienes describen la identificación de conformaciones “*active site*” y “*close conformations*”, según una definición subjetiva razonable. No sucede lo mismo para el ligando SU4984 dentro de 1agw, para quienes estos autores describen la identificación de una conformación próxima (“*close*”), mientras que en este caso no se ha podido posicionar correctamente con GOLDScore. De hecho, únicamente CHEMscore rinde resultados satisfactorios para este complejo. Pese a que AUTODOCK tampoco reproduce correctamente el modo de unión para la conformación de menor energía, sí que se recuperan conformaciones con una RMSD de 4 Å para SU4984 en 1agw que retienen las interacciones de puente de hidrógeno clave. Pese a que estas conformaciones no corresponden a las primeras del ranking, su *cluster* conformacional es uno de los más poblados. Este aspecto del análisis de la población de los *clusters* se discute más adelante.

Por otra parte, los resultados de *cross-docking* indican que la predicción de modos de unión fiables es ciertamente peor, especialmente en el caso de los resultados de AUTODOCK. GOLD es capaz de orientar SU5402 en 2fgi (1.1 Å) y 1agw (1.2 Å) con la función CHEMscore, e inversamente, SU4984 en 1fgi con GOLDScore (2.8 Å) y CHEMscore (1.7 Å). Curiosamente, en general para los cuatro protocolos, SU4984 se orienta y posiciona mejor en otra estructura (1fgi) que en la “suya” propia en la que fue cristalizado (1agw). En este sentido, existe una cierta concordancia con las conclusiones descritas en un estudio realizado por Diller y colaboradores<sup>381</sup>, quienes también encontraron problemas con la estructura 1agw. En su



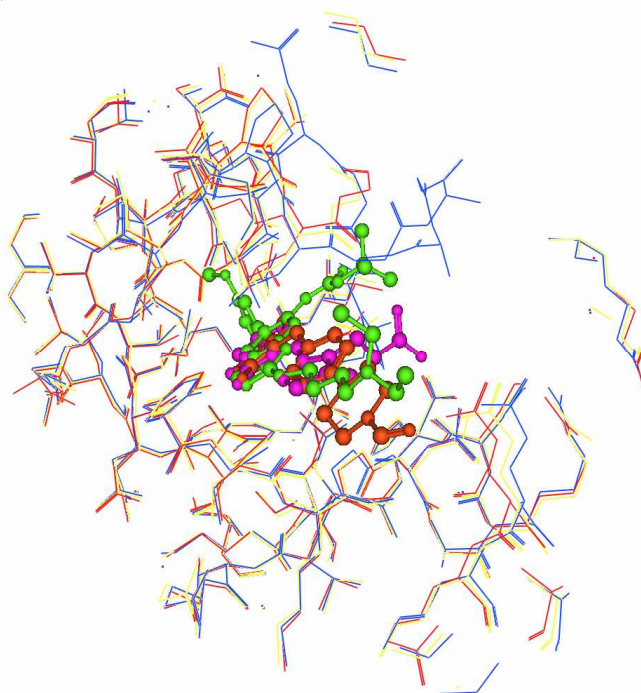
estudio, estos autores no fueron capaces de orientar una serie de ligandos conocidos en dicha estructura, alcanzando factores de enriquecimiento en un cribado virtual inferiores a una selección aleatoria. En su caso, un modelo de FGFR construido por homología utilizando la estructura de la quinasa Lck como plantilla, rindió resultados mejores que el uso de la estructura 1agw. Así, parece que 1agw se adapta excesivamente al ligando SU4984 para ajustarlo a su cavidad. Por otra parte, el tamaño pequeño de las indolinonas favorecería su *cross-docking* en estructuras más promiscuas como 2fgi.

En AUTODOCK, el tamaño de la malla fijado es relativamente grande, englobando la totalidad de la proteína. Así, se repite el *docking* con una malla de 101×101×101 para las posiciones de los ligandos nativos inicialmente situados en el sitio de unión y para los *cross-decoys*. No se observan diferencias en los valores de RMSD obtenidos (Tabla 5.2), respecto a los mostrados en la Tabla 5.1 calculados con mallas de dimensiones 181×181×181.

**Tabla 5.2.** Resultados de RMSD (Å) y sus desviaciones estándar (paréntesis) para las posiciones inicialmente situadas en el sitio de unión en el *docking* y el *cross-docking* frente a las tres estructuras resueltas de FGFR para los dos protocolos de AUTODOCK con tamaño de malla de 101×101×101.

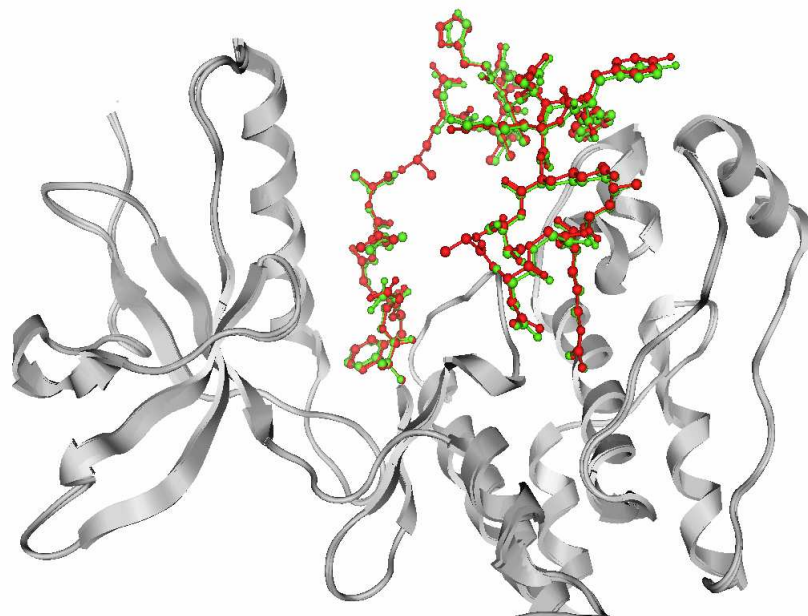
		PD173074	SU5402	SU4984
AUTODOCK	2fgi	<b>1.7 (0.1)</b>	15.2 (0.6)	7.0 (5.9)
Tipos atómicos estándar	1fgi	7.8 (2.3)	<b>2.4 (0.1)</b>	5.0 (0.3)
Cargas Gasteiger-Marsili	1agw	10.7 (3.5)	15.0 (0.2)	<b>9.8 (6.1)</b>
AUTODOCK	2fgi	<b>1.8 (0.2)</b>	16.1 (0.6)	9.7 (7.3)
Mapas de afinidad adicionales	1fgi	6.7 (2.2)	<b>2.6 (0.1)</b>	5.9 (0.9)
Cargas AM1	1agw	6.0 (2.9)	10.5 (5.3)	<b>12.1 (5.4)</b>

Para determinar porqué es más difícil acomodar el ligando PD1703074 en 1fgi y 1agw, se han comparado los *binding pockets* de las tres estructuras. En la Figura 5.2 se muestra la superposición de los sitios activos de 2fgi (*rojo*), 1fgi (*azul*) y 1agw (*amarillo*) dentro de un radio de 10 Å a partir de los inhibidores (mostrados en *ball and stick*). No se observan diferencias apreciables entre 2fgi y 1agw, y su RMSD correspondiente de superposición es inferior a 1 Å.



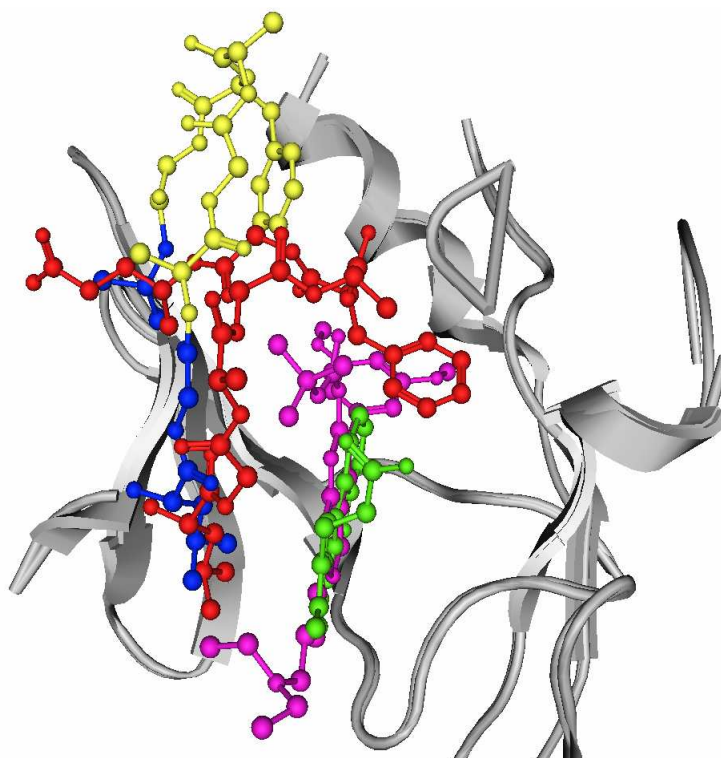
**Figura 5.2.** Superposición de los sitios activos de 2fgi (*rojo*), 1fgi (*azul*) y 1agw (*amarillo*) dentro de un radio de 10 Å a partir de los inhibidores (*ball and stick*)

Sin embargo, en muchos casos (63%), se ha observado que las conformaciones de PD173074 en las primeras posiciones del ranking se unen a una región particular de la estructura 1agw. Concretamente en la cavidad formada por los residuos Ala<sup>640</sup>-Trp<sup>665</sup>, correspondientes al dominio del *loop* de activación, comprendido entre los residuos 640 y 662. Por lo tanto, se ha comparado también esta región específica entre 1agw y 2fgi. Excepto por unos residuos no resueltos en la entrada 1agw (Arg<sup>646</sup>, Thr<sup>658</sup>), no se observa ninguna diferencia conformacional importante entre ambas estructuras (Figura 5.3).



**Figura 5.3.** Superposición de 1agw (*verde*) frente a 2fgi (*rojo*). Motivo del *loop* de activación donde muchas conformaciones de PD173074 tienen tendencia a unirse.

Sin embargo, sí que existe una diferencia entre el sitio de unión resuelto estructuralmente para 1fgi y 2fgi, ya que se encuentran cinco residuos resueltos en 1fgi que no están presentes en 2fgi (desde Glu<sup>486</sup> a Gly<sup>490</sup>). Debido a su proximidad al sitio de unión, estos residuos podrían afectar considerablemente al *docking*. De hecho, la posición del residuo Phe<sup>489</sup> difiere 10 Å entre 1fgi y 1agw, con su el anillo de fenilo próximo a SU5402. Por lo tanto, se decide modelar este fragmento para 2fgi transfiriendo las coordenadas desde 1fgi con el programa MODELLER<sup>131</sup>. En adelante, se refiere a esta estructura como 2fgi-transfer. En la Figura 5.4 se muestra el segmento que falta en 2fgi (*azul*) superpuesto a 1fgi (*rojo*) y 2fgi tras la transferencia de coordenadas (*amarillo*). Se observa cómo este pequeño motivo encierra al ligando SU5402 (*verde*) en su cavidad para 1fgi, mientras que se encuentra en una conformación más abierta en 2fgi y 2fgi-transfer.

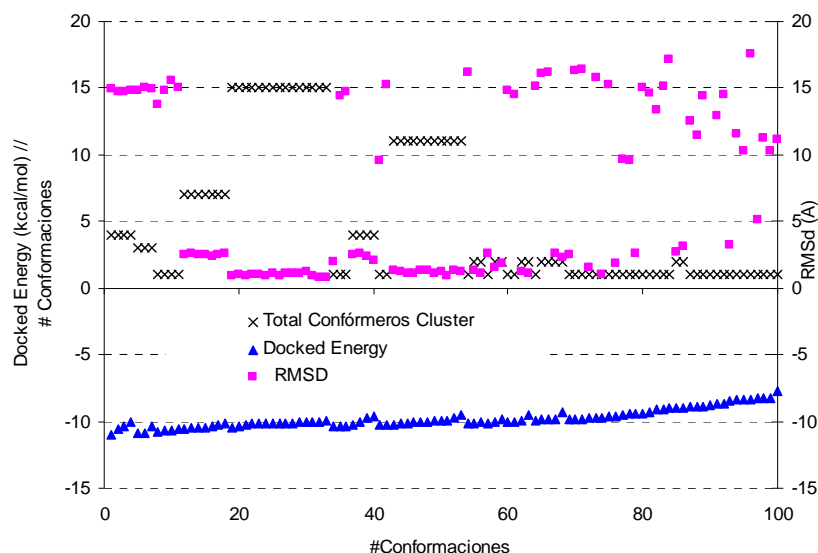


**Figura 5.4.** Detalle de los residuos Glu<sup>486</sup>-Gly<sup>490</sup> no presentes en 2fgi (*azul*) comparados con 1fgi (*rojo*) y 2fgi tras la transferencia de coordenadas (*amarillo*). SU5402 se muestra en verde y PD173074 en magenta.

Tras repetir los diferentes *dockings* frente a la estructura 2fgi-transfer (Tabla 5.1), no se observa una diferencia significativa comparada con los resultados de RMSD obtenidos frente a 2fgi, excepto en el caso de la función CHEMSCORE, donde empeora. Por lo tanto, la transferencia de coordenadas no es suficiente para reproducir el modo de unión de 1fgi en 2fgi.

Los resultados mostrados en la Tabla 5.1 corresponden a la conformación en la primera posición del ranking según su *scoring* de los 100 conformeros requeridos en cada *docking*. Existe la posibilidad de que los resultados fallidos en la reproducción del modo de unión sean debidos al procedimiento de ranking, es decir, que se identifique correctamente el modo de unión pero que dicha conformación no ocupe la posición prioritaria del ranking. De hecho, se ha observado en diversos *dockings* que la conformación correcta, aunque no correctamente ordenada en el ranking, pertenece a uno de los *clusters* que presentan un mayor número de conformeros (se agrupan en un *cluster* aquellos con una RMSD de 1.0 Å).

En la Figura 5.5 se ilustra un ejemplo de esta situación: para el *docking* de SU5402 en 2fgi con AUTODOCK. Los 100 conformeros resultantes aparecen agrupados por *clusters*, estableciéndose el orden relativo de los *clusters* de mayor a menor energía de *docking* según el valor que presenta el conformero con mayor *docked energy* de cada *cluster*. El *cluster* más poblado contiene 15 conformaciones, con una RMSD promedio respecto a la conformación cristalizada de 0.99 Å, por lo que este *cluster* identifica con éxito el modo de unión. Sin embargo, su energía final de *docking* (*docked energy* en AUTODOCK) es ligeramente superior a la de la conformación de mínima energía, posicionada a 15 Å del *binding pocket* y que pertenece a un *cluster* con únicamente 4 conformaciones. Se desea enfatizar el hecho de que AUTODOCK devuelve dos energías: la denominada *docked energy*, utilizada para guiar el *docking*, y la energía libre de unión (*binding free energy*). En la Figura 5.5 únicamente se muestra el valor de la *docked energy*, valor que no debe interpretarse como una medida de la afinidad de la unión.



**Figura 5.5.** RMSD (Å), *Docked energy* (kcal/mol) y número de conformaciones en un *cluster* para las 100 conformaciones requeridas en el *docking* de SU5402 en 2fgi con AUTODOCK y el protocolo de cargas Gasteiger-Marsili.

#### 5.1.4. Resultados del *docking* ciego en FGFR

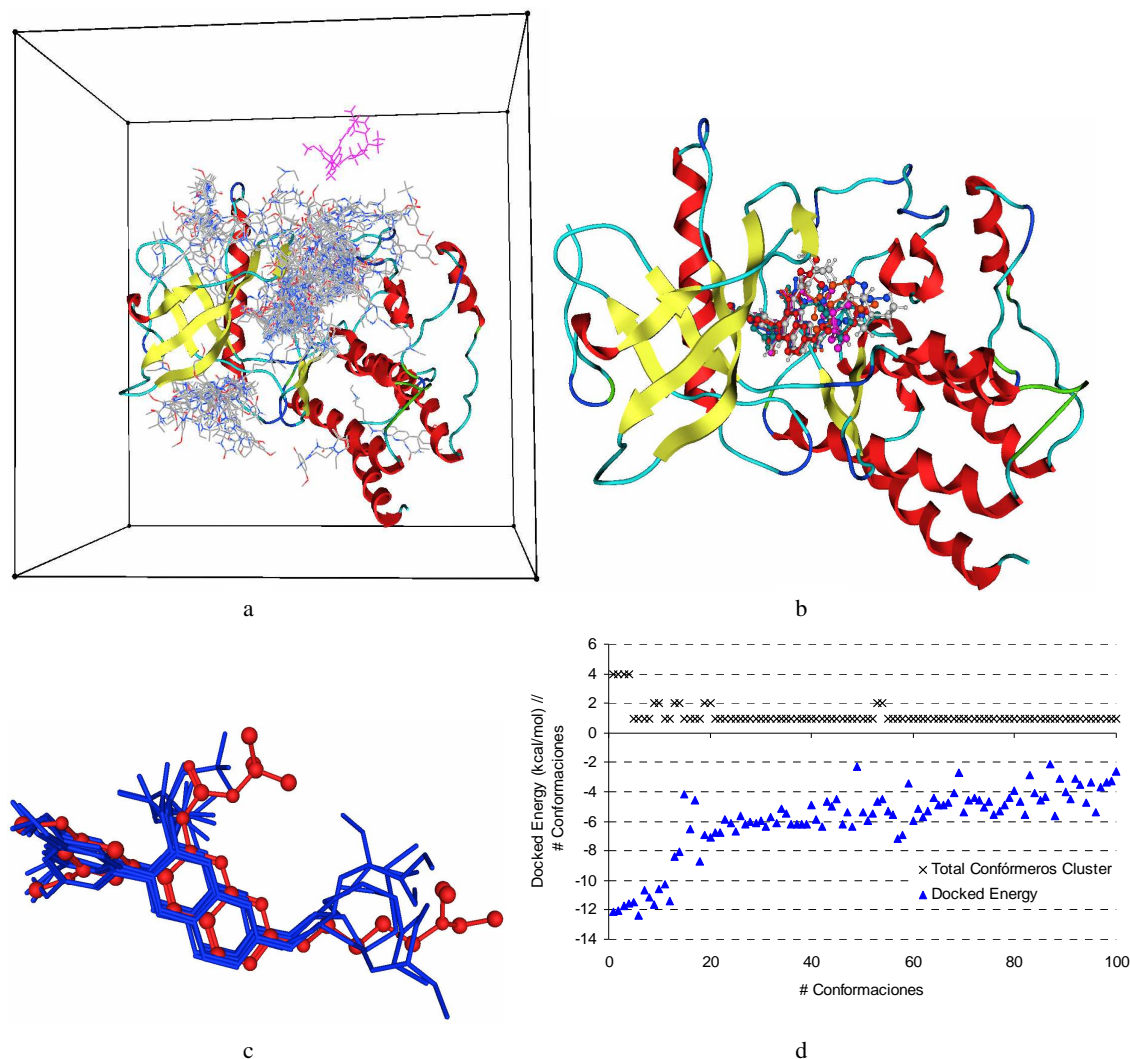
Los resultados de RMSD para el *docking* ciego de PD173074, SU5402 y SU4984 en sus respectivas estructuras, junto a los resultados para 2fgi-transfer se muestran en la Tabla 5.3 para los cálculos con AUTODOCK. También se incluyen los resultados para el *cross-docking* ciego.

**Tabla 5.3.** Resultados de RMSD (Å) y sus desviaciones estándar (paréntesis) para las posiciones inicialmente situadas en la superficie de la proteína en el *docking* ciego y el *cross-docking* ciego frente a las tres estructuras resueltas de FGFR para los dos protocolos de AUTODOCK.

		PD173074	SU5402	SU4984
AUTODOCK Tipos atómicos estándar Cargas Gasteiger-Marsili	2fgi	<b>1.8 (0.1)</b>	15.4 (0.6)	12.4 (8.1)
	2fgi-transfer	1.8 (0.1)	15.0 (0.4)	14.1 (4.9)
	1fgi	10.2 (1.7)	<b>2.5 (0.2)</b>	5.1 (0.5)
	1agw	9.2 (4.3)	15.4 (0.2)	<b>11.9 (6.1)</b>
AUTODOCK Mapas de afinidad adicionales Cargas AM1	2fgi	<b>1.7 (0.1)</b>	15.2 (0.4)	12.6 (6.9)
	2fgi-transfer	1.7 (0.4)	15.1 (0.2)	17.1 (1.7)
	1fgi	9.5 (4.1)	<b>2.5 (0.2)</b>	6.1 (0.3)
	1agw	10.4 (3.7)	15.6 (0.4)	<b>13.1 (6.4)</b>

Es destacable cómo para los casos de los ligandos frente a sus estructuras nativas, el procedimiento de *docking* ciego es capaz de identificar correctamente el sitio y el modo de unión para 1fgi y 2fgi.

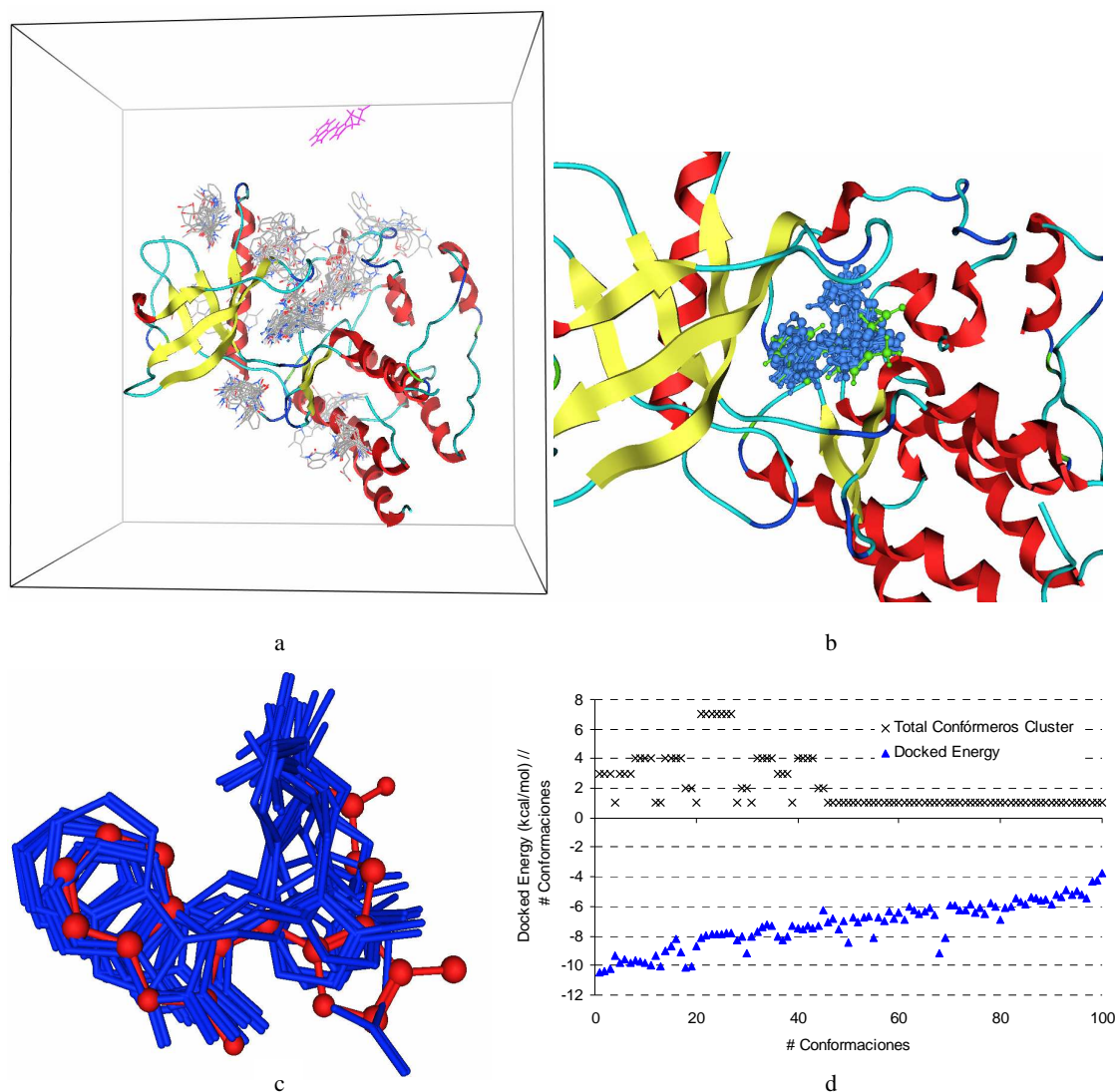
En la Figura 5.6 se muestran los resultados obtenidos para el *docking* ciego de PD173074 frente a 2fgi. De las 100 conformaciones obtenidas (Figura 5.6.a), las de menor *docked energy* encuentran perfectamente la cavidad (Figura 5.6.b), mostrándose su superposición frente a PD173074 co-cristalizado, en rojo en la Figura 5.6.c. En la Figura 5.6.a la conformación en rosa corresponde a la posición inicial del ligando. Asimismo, se muestra la distribución en *clusters* de los 100 conformeros calculados (Figura 5.6.d), observándose cómo aquellos de menor *docked energy*, que concuerdan con los de menor RMSD de superposición, pertenecen a los *clusters* más poblados.



**Figura 5.6.** Resultados del *docking* ciego para PD173074 frente a 2fgi con AUTODOCK y el protocolo con cargas Gasteiger-Marsili. (a) 100 conformaciones resultantes. (b) conformaciones de menor *docked energy* que encuentran el sitio activo. (c) superposición de las primeras conformaciones del ranking (azul) frente a PD173074 co-cristalizado (rojo). (d) Distribución de las 100 conformaciones en *clusters* y su energía de *docking* respectiva (*docked energy*).

En la Figura 5.7 se muestran los resultados obtenidos para el *docking* ciego de SU5402 frente a 1fgi. En la Figura 5.7.a se muestran las 100 conformaciones obtenidas y la conformación inicial de partida (rosa), donde las 12 primeras del ranking corresponden a conforméromos que detectan correctamente el sitio de unión de SU5402 (conforméromos en azul en la Figura 5.7.b). La superposición de estas conformaciones frente a SU5402 co-cristalizado (rojo) se detalla en la Figura 5.7.c. Siguiendo con el análisis de *clusters* (Figura 5.7.d), estas primeras 12 conformaciones corresponden tanto a las de mínima energía de *docking* como a *clusters* con varios representantes. Sin embargo, en este caso el análisis de población de *clusters* por sí sólo no concuerda totalmente con las mejores conformaciones, ya que el *cluster* más poblado (con 7 representantes) corresponde a conforméromos posicionados fuera de la cavidad de unión. En este caso, el criterio energético o de *scoring* es el que permite detectar la conformación que mejor reproduce el modo (y sitio) de unión.



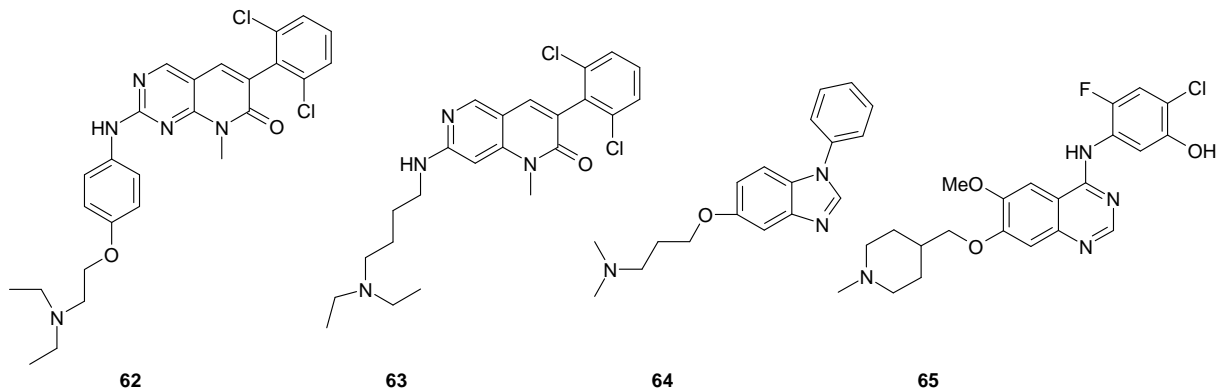


**Figura 5.7.** Resultados del *docking* ciego para SU5402 frente a 1fgi con AUTODOCK y el protocolo con cargas Gasteiger-Marsili. (a) 100 conformaciones resultantes, en *rosa* la conformación de partida (b) conformaciones de menor *docked energy* que encuentran el sitio activo. (c) superposición de las primeras conformaciones del ranking (*azul*) frente a SU5402 co-cristalizado (*rojo*). (d) Distribución de las 100 conformaciones en los *clusters* y su energía de *docking* respectiva (*docked energy*).

Observando los valores de RMSD obtenidos para los *cross-decoys* con AUTODOCK en *docking* ciego (Tabla 5.3), se aprecia como tampoco es posible reproducir el modo y sitio de unión para los ligandos intercambiados en las otras estructuras. La estructura 2fgi-transfer se comporta de nuevo análogamente a la estructura 2fgi, y 1agw no permite la reproducción del modo de unión de su “propio” ligando (SU4984). Basándose en estos resultados y en los mostrados en el apartado anterior, se decide continuar el estudio considerando únicamente las estructuras 1fgi y 2fgi.

Se ha realizado también un *docking* ciego para un representante de cada una de las otras cuatro series químicas de compuestos activos presentes en el *pool* ACTIV\_1 (apartado 4.1): pirido[2,3-*d*]pirimidinas, naftiridin-2(1*H*)-onas, 1-fenilbenzimidazoles y 4-anilinoquinazolininas. Estos cuatro compuestos corresponden a los seleccionados en el apartado 4.2 como plantillas para la generación de modelos farmacofóricos (Figura 5.8).

Sus estructuras se construyen en MOE, se colocan aleatoriamente en la superficie de la proteína y se optimiza su geometría con el *force field* MMFF94 y el procedimiento por defecto de MOE, con un gradiente de convergencia de RMS de 0.05. Para cada uno de ellos, se han aplicado los dos protocolos de AUTODOCK descritos en el apartado 5.1.2: con los tipos atómicos estándar y las cargas Gasteiger-Marsili y con mapas de afinidad adicionales y cargas AM1, tanto frente a 1fgi como a 2fgi.



**Figura 5.8.** Cuatro compuestos representativos de cada una de las series químicas del *pool* ACTIV\_1 seleccionados para el *docking* ciego.

Dado que la estructura cristalina de estos ligandos en FGFR no ha sido resuelta, no es factible medir una RMSD para evaluar las conformaciones obtenidas. Por lo tanto, las soluciones se inspeccionan visualmente y se clasifican en dos grupos: *active site* (A) y *misdoocked* (M). La clasificación *active site* incluye aquellas predicciones donde se encuentra el sitio de unión y la conformación se superpone con el correspondiente inhibidor co-cristalizado (PD173074 o SU5402) del mismo modo a como se superpone con la herramienta de alineamiento flexible de MOE (apartado 4.3), manteniéndose las interacciones por puente de hidrógeno clave. La asignación *misdoocked* se refiere a conformaciones en las que las dos condiciones no se cumplen, aunque principalmente se debe a casos en los que el algoritmo no es capaz de detectar apropiadamente el sitio de unión. Los resultados se resumen en la Tabla 5.4.

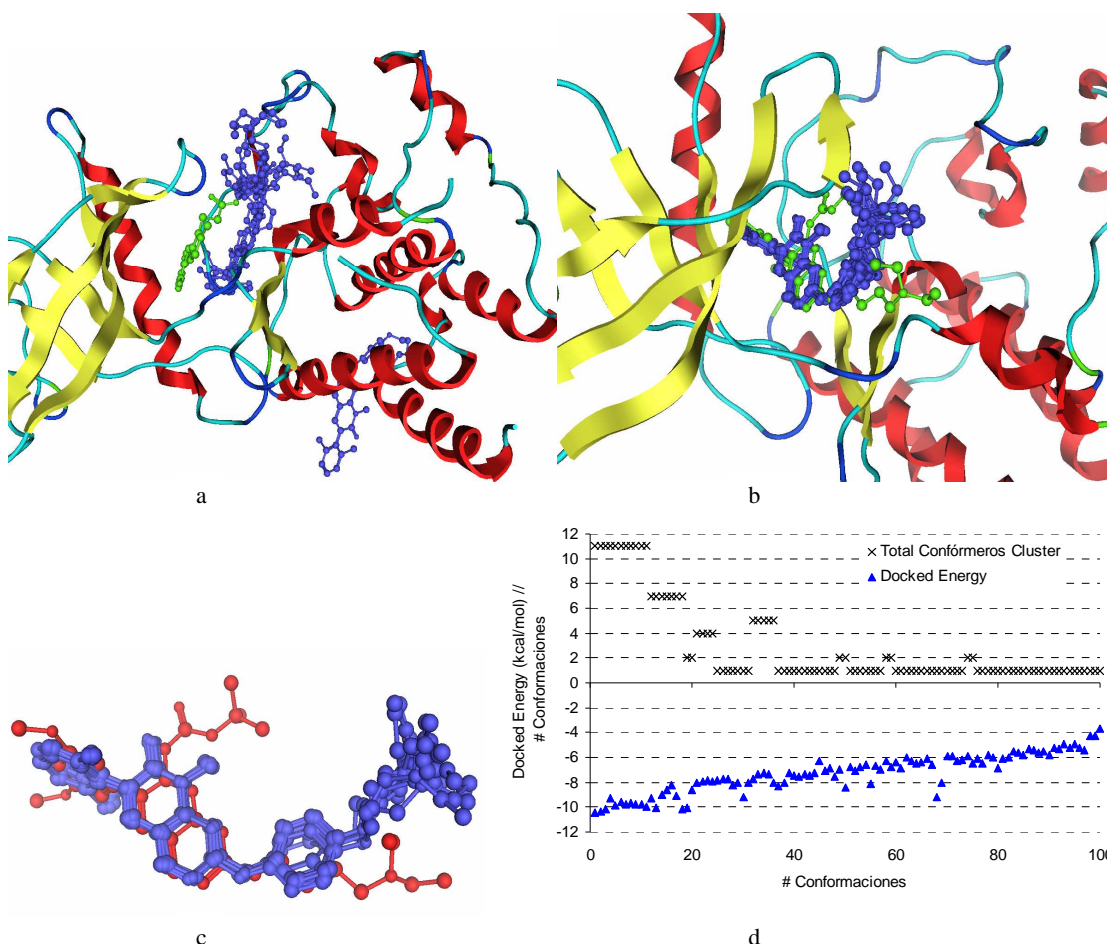
**Tabla 5.4.** Resumen de las predicciones del *docking* ciego frente a FGFR para cuatro compuestos activos diferentes de los ligandos co-cristalizados.

	AUTODOCK		AUTODOCK	
	Tipos atómicos estándar		Mapas de afinidad adicionales	
	Cargas Gasteiger-Marsili		Cargas AM1	
	1fgi	2fgi	1fgi	2fgi
62	M	A	M	A
63	M	A	M	A
64	A	A <sup>a</sup>	A	M
65	M	M	M	M

<sup>a</sup> corresponde a un caso en el que la conformación *active site* se encuentra en un cluster muy poblado, aunque queda en la segunda posición del ranking, tras una conformación *misdoocked*.

Se observa claramente cómo el quimiotipo del ligando co-cristalizado influencia el tipo de *scaffold* activo reconocido. La pirido[2,3-*d*]pirimidina (**62**) y la naftiridin-2(1*H*)-ona (**63**) se orientan correctamente en 2fgi, mientras que el 1-fenilbenzimidazol (**64**) lo hace preferentemente en 1fgi. Como en el caso de los *cross-decoys*, 1fgi fracasa en el reconocimiento de los heterociclos con seis anillos. Para la 4-anilinoquinazolina (**65**), las mejores conformaciones del ranking se posicionan fuera del sitio de unión, tanto para 1fgi como para 2fgi.

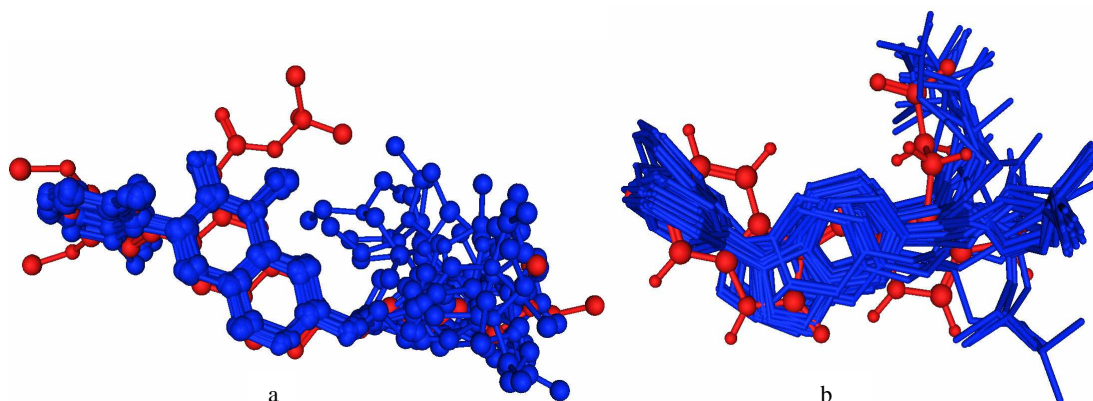
En la Figura 5.9 se ejemplifican las dos posibles soluciones (*active site* y *misdocked*) para el *docking* de la pirido[2,3-*d*]pirimidina (**62**) en 1fgi (Figura 5.9.a) y 2 fgi (Figura 5.9.b) con AUTODOCK y el protocolo con cargas Gasteiger-Marsili. El inhibidor co-cristalizado correspondiente (SU5402 o PD173074, respectivamente) se muestra en verde y las mejores conformaciones del *docking* aparecen en azul. En el caso de 1fgi los resultados están totalmente *misdocked* mientras que la predicción en 2fgi reproduce con éxito el modo de unión. En la Figura 5.9.c, se ilustra un *zoom* de la superposición del conjunto de conformaciones en las primeras posiciones del ranking (*azul*) frente a PD173074 (*rojo*). Excepto para la cadena flexible, descrita como de difícil resolución estructural<sup>371</sup>, para el resto de la estructura se observa una buena superposición. Aquí, la posición prioritaria del ranking corresponde al *cluster* más poblado de conformaciones (11), por lo que en este caso existe un acuerdo entre el criterio de *scoring* y el número de conformaciones en el *cluster* (Figura 5.9.d).



**Figura 5.9.** Resultados del *docking* ciego para el compuesto **62**, representativo de la serie pirido[2,3-*d*]pirimidina frente a (a) 1fgi y (b) 2fgi con AUTODOCK y el protocolo con cargas Gasteiger-Marsili. Los residuos en *azul* corresponden a las conformaciones predichas y el *verde* corresponde al inhibidor co-cristalizado en cada caso. (c) Superposición de las primeras conformaciones del ranking (*azul*) frente a PD173074 co-cristalizado (*rojo*). (d) Distribución de las 100 conformaciones en los *clusters* y su energía de *docking* respectiva (*docked energy*).



Finalmente, se muestran las conformaciones clasificadas como *active site* encontradas en el *docking* ciego de la naftiridin-2(1*H*)-ona (**63**) frente a 2fgi, superpuestas al PD173074 co-cristalizado (Figura 5.10.a) y del 1-fenilbenzimidazol (**64**) frente a 1fgi, superpuestas frente a SU5402 co-cristalizado (Figura 5.10.b). Exceptuando las cadenas laterales, la superposición es satisfactoria. Además, coincide con los resultados obtenidos mediante el alineamiento flexible, presentados en el apartado 4.3.



**Figura 5.10.** (a) 18 primeras conformaciones de la naftiridin-2(1*H*)-ona (**63**) (azul) en el ranking del *docking* ciego frente a 2fgi superpuestas sobre el PD173074 co-cristalizado (rojo). (b) 45 primeras conformaciones del 1-fenilbenzimidazol (**64**) (azul) en el ranking del *docking* ciego frente a 1fgi superpuestas frente a SU5402 co-cristalizado (rojo).

### 5.1.5. Cribado virtual en FGFR

Para la validación retrospectiva del cribado frente a FGFR, se selecciona la base de datos Base\_ACTIV\_1 (apartado 4.1), formada por 288 inhibidores de FGFR y 4931 compuestos inactivos. El *pool* de inactivos presenta unas propiedades 1D similares a las de los compuestos activos, lo que permite evitar una posible desviación de los resultados del *docking* generada por las diferencias en tamaño molecular. Este tipo de diseño inicial de la base de datos, ha sido denominado *focused libraries* por Verdonk y colaboradores<sup>215</sup>, los cuales han demostrado la importancia de evitar este tipo de desviaciones entre los compuestos activos y los inactivos. En la Tabla 5.5 se muestran el promedio y la desviación estándar (en paréntesis) del peso molecular, número de enlaces rotables simples (b\_1rotN), número de aceptores de puente de hidrógeno (a\_acc) y número de átomos dadores (a\_don), calculados con MOE para los dos *pools*. Examinando dicha tabla, se concluye que no existen grandes diferencias apreciables entre el *pool* de activos e inactivos.

**Tabla 5.5.** Propiedades 1D del *pool* de activos e inactivos de la base de datos Base\_ACTIV\_1. En paréntesis se muestra la desviación estándar.

	Peso Molecular	b_1rotN	a_acc	a_don
288 Activos	416.0 (146.4)	8.5 (4.0)	3.6 (1.3)	1.9 (1.3)
4931 Inactivos	415.7 (95.5)	10.6 (6.5)	3.9 (2.5)	2.4 (2.2)

El cribado virtual frente a FGFR se realiza con los programas AUTODOCK y GOLD para las dos estructuras 1fgi y 2fgi. A continuación se describen los detalles de la metodología seguida con cada uno de ellos en el cribado virtual frente a FGFR. Lo mismo aplica para EGFR y PDGFR, excepto aspectos puntuales indicados en sus correspondientes apartados.

- **AUTODOCK.** A la vista de los resultados obtenidos en los apartados anteriores (Tablas 5.1, 5.3 y 5.4) no se encuentran diferencias significativas en las soluciones del *docking* entre los dos protocolos de AUTODOCK empleados. Por lo tanto, se opta por realizar el cribado virtual con el protocolo con las cargas Gasteiger-Marsili, que requiere menos tiempo computacional que MOPAC. Se incluyen los mapas de afinidad adicionales para tipos atómicos. Respecto a los parámetros expuestos en el apartado 5.1.2, se modifica el tamaño de la malla a 101×101×101, requiriéndose 10 confórmeros para cada compuesto de la base de datos.

En este trabajo se ha utilizado la versión 3.0 de AUTODOCK, sin incluir la herramienta gráfica AutoDockTools (ADT). Este programa no dispone de una aplicación para el tratamiento completo de una base de compuestos (típicamente en formato *sd-file*) por lo que se han programado diversas *scripts* que permitan la automatización del proceso de *docking*:

- ◆ Acondicionamiento de los ligandos con el programa MOE (lenguaje SVL)<sup>38</sup>:
  - *Script* Alinear.svl: Alinea las moléculas contenidas en una base de datos sobre el ligando co-cristalizado en la cavidad de unión. El objetivo es posicionar todas las moléculas inicialmente en dicho sitio activo, por lo que pese a que utiliza la función de alineamiento flexible de MOE, las opciones activadas en este caso por la *script* son muy poco restrictivas, a fin de que el proceso sea rápido. Como salida se obtiene una conformación (por defecto la de menor valor de función objetivo, S) para cada molécula, en formato *mol2*.
  - *Script* Cargar.svl: Permite asignar las cargas, según el método escogido (Gasteiger-Marsili...) todos los compuestos, guardándolos en formato *mol2*.
- ◆ *Docking* automático de todos los compuestos de una base de datos mediante una *script* en perl (*Practical Extraction and Report Language*). Se destacan dos características a tenerse en cuenta:
  - Generación automática del fichero en formato *.pdbq* para cada ligando mediante el código autotors de AUTODOCK que establece las torsiones permitidas para el ligando flexible. Este código ha sido modificado para que no se active el menú interactivo en modo texto, sino que se seleccione automáticamente como *root* en la asignación de torsiones posibles el primero de los ciclos identificados. También se aceptan por defecto todas las posibles torsiones propuestas por autotors.
  - Manipulación automática del fichero *.pdbq* obtenido para cambiar la nomenclatura de los tipos atómicos adicionales: bromos, cloros, nitrógenos dadores de puente de hidrógeno, etc, ya que dicha nomenclatura difiere de la comúnmente empleada en los paquetes de diseño molecular.
- ◆ Análisis de los resultados. AUTODOCK dispone de una *script* para extraer todas las conformaciones requeridas del ligando en formato *pdb*. Sin embargo, dicho formato contiene la nomenclatura de tipos atómicos establecida en AUTODOCK, por lo que los ficheros de salida no son perfectamente legibles por otros programas como MOE. Por lo tanto, se ha programado una *script* en perl que permite extraer todas o una conformación en particular (la de mínimo *scoring* o la contenida en el

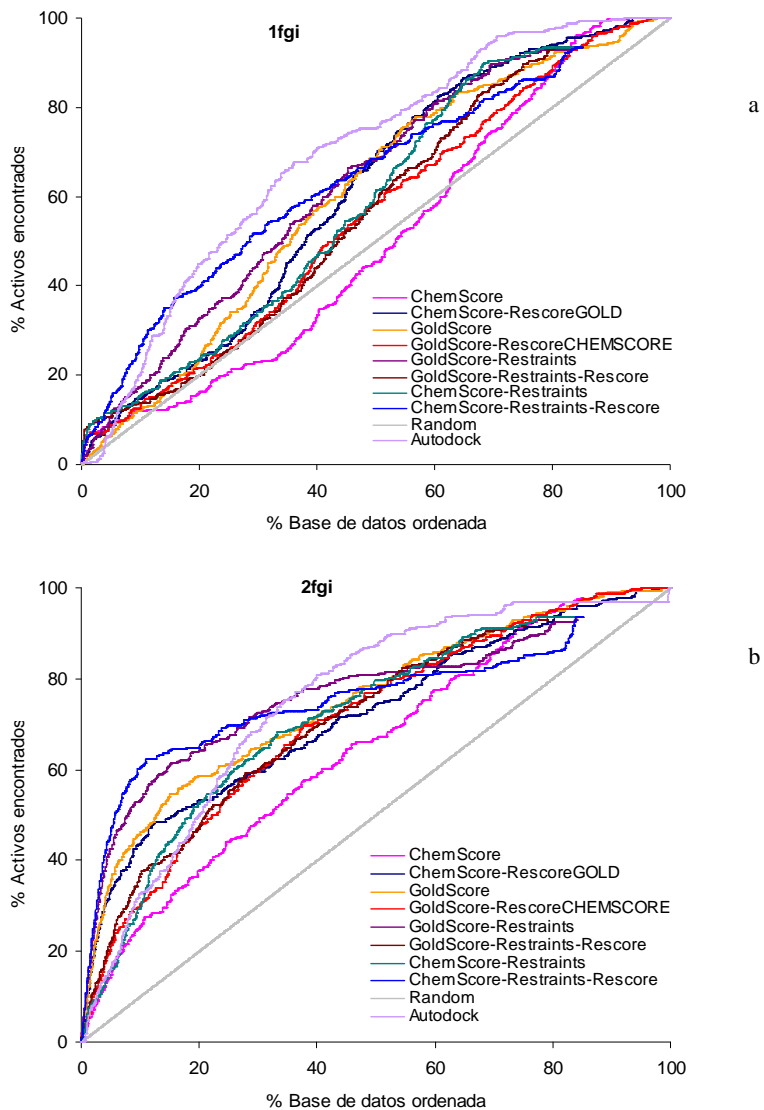
*cluster* más poblado), modificando su nomenclatura atómica. También extrae una tabla resumen para todos los conformeros y/o el conformero de interés (típicamente el de mayor valor absoluto de energía) conteniendo información de sus valores de *docked energy*, *binding free energy*, RMSD a una estructura de referencia y la población del *cluster* al que pertenece dicho conformero.

- **GOLD.** Este programa presenta la ventaja de trabajar directamente sobre una base de datos almacenada en formato *sd-file*. Para el cribado virtual, se ejecuta el programa con los parámetros por defecto optimizados para el cribado de quimiotecas (*library screening settings*), que pese a ser menos exhaustivos que los descritos en el apartado 5.1.2, se consideran lo suficientemente óptimos para esta finalidad, a la par que se reduce el tiempo de cálculo. Para cada ligando, se requieren 10 conformeros. En total, se analizan cuatro protocolos de *docking/scoring* con el programa GOLD:
  - ◆ GOLDScore: la misma función de *scoring* para guiar el *docking* y evaluar cada ligando
  - ◆ CHEMScore: la misma función de *scoring* para guiar el *docking* y evaluar cada ligando.
  - ◆ GOLDScore-REScore: Las soluciones del *docking* producidas con la función GOLDScore se re-evalúan (*rescoring*) con la función CHEMScore.
  - ◆ CHEMScore-REScore: Las soluciones del *docking* con la función CHEMScore se re-evalúan con la función GOLDScore.

Cada una de las cuatro situaciones se realiza en presencia o ausencia de una serie de restricciones (*constraints o restraints*). Éstas requieren que el ligando satisfaga las funcionalidades establecidas responsables del modo de unión del ATP, imponiendo *constraints* tipo “*Protein Hydrogen Bonds*” con un factor de peso de 10. En el caso del 1fgi, se requiere que los ligandos formen un puente de hidrógeno con el oxígeno del carbonilo del Glu<sup>562</sup> o con el nitrógeno amídico de la Ala<sup>564</sup>. En el caso de 2fgi, las dos posibles restricciones de puente de hidrógeno se asignan al nitrógeno amídico de la Ala<sup>564</sup> o al oxígeno del carbonilo del residuo Ala<sup>564</sup>. Siguiendo las indicaciones del programa, es necesario asegurarse de que GOLD incluya y considere dichas restricciones. Para acelerar el proceso, se prescinde del *docking* de aquellos ligandos para los que las restricciones son imposibles de satisfacer.

En resumen, en total se examinan un total de ocho protocolos diferentes con GOLD (cuatro×dos, con o sin restricciones).

En la Figura 5.11 se muestran las curvas de enriquecimiento obtenidas en el *docking* frente a 1fgi (a) y 2fgi (b). Estas curvas representan el porcentaje de activos conocidos recuperados *versus* el porcentaje de base de datos ordenada que ha sido cribada. El ranking de la base de datos se establece de acuerdo al valor de *scoring* del *docking*.



**Figura 5.11.** Curvas de enriquecimiento obtenidas con los nueve protocolos de *docking* frente FGFR para las estructuras (a) 1fgi y (b) 2fgi en el cribado de Base\_ACTIV\_1. La línea gris representa los valores obtenidos si los activos se seleccionasen aleatoriamente.

En la Tabla 5.6 se muestran los factores de enriquecimiento (*ef*), véase apartado 4.4, en el primer 2%, 5% y 10% de base de datos Base\_ACTIV\_1 cribada.

**Tabla 5.6.** Factores de enriquecimiento (*ef*) obtenidos en el primer 2%, 5% y 10% de base de datos cribada frente a FGFR.

	2%		5%		10%	
	1fgi	2fgi	1fgi	2fgi	1fgi	2fgi
CHEMSCORE	3.6	3.3	1.8	3.0	1.2	2.5
CHEMSCORE-RESCORE	2.6	9.3	1.8	6.7	1.5	4.3
CHEMSCORE-FILTRADO	3.7	3.8	1.8	3.3	1.0	2.7
CHEMSCORE-RESTRAINT	4.8	4.5	2.4	8.5	1.5	3.0
CHEMSCORE-RESTRAINT-RESCORE	3.8	12.4	2.8	4.4	2.7	6.1
GOLDScore	1.4	9.0	1.3	7.1	1.2	4.6
GOLDScore-RESCORE	3.5	4.8	1.7	4.2	1.3	3.0
GOLDScore-FILTRADO	1.5	9.3	1.5	7.0	1.2	4.4
GOLDScore-RESTRAINT	2.4	11.4	2.2	8.5	1.7	5.3
GOLDScore-RESTRAINT-RESCORE	4.7	5.4	2.2	4.4	1.4	3.7
AUTODOCK	0.3	4.0	1.4	3.2	2.0	3.3

El cribado virtual frente a 2fgi supera a los resultados del *docking* con 1fgi en todos los casos, mostrando cómo la estructura de la proteína afecta significativamente los factores de enriquecimiento obtenidos. Para 1fgi, muchos protocolos no están lejos de ser una selección aleatoria o incluso peor, como es el caso del cribado con AUTODOCK en el primer 2% de base de datos o la función CHEMSCORE a porcentajes más elevados de base de datos. En el primer 10% de base de datos ordenada, muchas de las funciones de *scoring* rinden valores de *ef* próximos a la unidad para el *docking* frente a 1fgi.

Aparte de consideraciones sobre la estructura de la proteína, se observa cómo la inclusión de información estructural acerca del modo de unión en la forma de *constraints*, mejora el enriquecimiento. Particularmente, en el caso de 2fgi, los valores de *ef* en el primer 2% de base de datos se incrementan desde 3.3 a 4.5 (CHEMSCORE) y de 9.0 a 11.4 (GOLDScore). La inclusión de *constraints* implica dos factores. En primer lugar, aquellos compuestos que no satisfacen el motivo del puente de hidrógeno son automáticamente excluidos. En segundo lugar, se incluye un término adicional de *scoring* en la función de *scoring* (apartado 1.4.1). Para comprobar hasta qué punto cada uno de estos dos factores es responsable de la mejora en el enriquecimiento, se filtran manualmente aquellos compuestos que no satisfacen el motivo del puente de hidrógeno de la base de datos cribada sin *constraints*. A continuación, se recalculan los factores de enriquecimiento. Estos valores se muestran en la Tabla 5.6, bajo las etiquetas “CHEMSCORE-FILTRADO” y “GOLDScore-FILTRADO”. Para 2fgi, el valor de *ef* en el primer 2% de la base de datos cribada se incrementa únicamente desde 3.3 hasta 3.8 (CHEMSCORE) y desde 9.0 hasta 9.3 (GOLDScore). Así, en ambos casos, el filtrado manual no es suficiente para reproducir la situación con *constraints*, con valores de *ef* de 4.5 y 11.4, respectivamente. Aunque la etapa de filtrado mejora ligeramente los resultados de la situación sin *constraints*, se deduce que el principal responsable de la mejora del enriquecimiento alcanzada con la inclusión de *constraints* es el término adicional impuesto en la función de *scoring*.

Por otra parte, el *rescoring* con GOLDScore es útil para discriminar entre compuestos activos e inactivos, rindiendo el valor máximo de *ef* de 12.4 para 2fgi en el primer 2% de la base de datos cribada. El valor máximo teórico del factor de enriquecimiento para Base\_Activ\_1 es de 18.1% (5219/288).

De manera análoga a como se ha realizado en la aproximación *ligand-based* (capítulo 4), se analiza también en este caso la capacidad de las estructuras de FGFR (y de los algoritmos de *docking*) de recuperar un *pool* diverso de *scaffolds* que pudiera facilitar la identificación de compuestos *lead*. Así, se determina el porcentaje de base de datos que tiene que ser cribada para que al menos un compuesto activo representativo de cada *scaffold* activo en Base\_ACTIV\_1 sea recuperado (Tabla 5.7).

Otros autores han apuntado previamente a la importancia de la identificación de quimiotipos diversos como una medida de la robustez de un algoritmo de *docking*.<sup>389</sup>

Para cada uno de los compuestos representativos recuperados, se evalúa visualmente si la conformación obtenida reproduce o no el modo de unión previsible, siguiendo la clasificación del apartado 5.1.4 en *active site* (A) o *misdocked* (M). Además, en la Tabla 5.7 se indica la existencia (1) o no (0) de un puente de hidrógeno entre el correspondiente compuesto y los residuos clave de FGFR. Nótese que la identidad del compuesto representante de cada *scaffold* puede diferir o no entre los distintos protocolos, según sean ordenados todos los compuestos de la base de datos por su *scoring*, aunque se trate de la misma conformación (caso de aplicar *rescoring*).

**Tabla 5.7.** Porcentaje de base de datos cribada frente a FGFR para la que al menos un representante activo de cada uno de los seis *scaffolds* de Base\_Activ\_1 es recuperado.

	1fgi						2fgi					
	PD	Naph	Urea	Indol	Phen	Anil	PD	Naph	Urea	Indol	Phen	Anil
CHEMSCORE	<b>35.1</b>	14.8	19.9	0.1	6.0	25.4	1.4	0.1	14.5	0.1	6.8	<b>29.7</b>
	M(0)	M(0)	M(0)	A(1)	A(1)	M(1)	A(1)	A(1)	A(1)	A(1)	M(0)	M(0)
CHEMSCORE-RESCORE	4.9	3.8	1.5	0.1	21.3	<b>28.8</b>	0.2	1.4	0.1	0.1	17.1	<b>20.0</b>
	M(0)	M(1)	M(0)	A(1)	M(0)	M(0)	A(1)	A(1)	A(1)	A(1)	M(0)	M(0)
CHEMSCORE-RESTRAINT	<b>17.8</b>	8.6	3.9	0.1	9.4	11.9	1.6	0.8	2.1	0.1	7.3	<b>29.5</b>
	M(1)	M(1)	M(1)	A(1)	A(1)	M(1)	A(1)	A(1)	A(1)	A(1)	M(1)	M(0)
CHEMSCORE-RESTRAINT-RESCORE	2.8	3.8	0.1	0.1	27.1	<b>27.3</b>	0.6	1.0	0.1	0.1	14.8	<b>65.3</b>
	M(1)	M(1)	M(1)	A(1)	A(1)	M(0)	A(1)	A(1)	A(1)	A(1)	M(1)	M(0)
GOLDSCORE	1.8	18.8	12.8	0.3	10.2	<b>67.4</b>	0.2	2.6	0.1	0.4	9.6	<b>32.1</b>
	M(0)	M(0)	M(0)	A(1)	M(0)	M(0)	A(1)	A(1)	A(1)	A(1)	M(0)	M(1)
GOLDSCORE-RESCORE	14.8	5.8	25.0	0.1	3.8	<b>26.6</b>	0.7	0.5	1.3	0.1	2.8	<b>26.6</b>
	M(0)	M(0)	M(0)	A(1)	A(1)	M(0)	A(1)	A(1)	A(1)	A(1)	M(0)	M(0)
GOLDSCORE-RESTRAINT	1.9	<b>35.4</b>	18.3	0.1	17.7	<b>35.4</b>	0.4	3.3	0.1	0.1	18.2	<b>36.2</b>
	M(1)	M(1)	M(1)	A(1)	A(1)	M(1)	A(1)	A(1)	A(1)	A(1)	M(1)	M(0)
GOLDSCORE-RESTRAINT-RESCORE	8.5	5.9	2.9	0.1	9.5	<b>23.9</b>	1.1	1.0	0.9	0.0	5.7	<b>18.9</b>
	M(1)	M(1)	M(1)	A(1)	A(1)	M(1)	A(1)	A(1)	A(1)	A(1)	M(1)	M(0)
AUTODOCK	3.8	15.4	1.6	3.9	<b>32.2</b>	0.0	0.6	0.7	0.7	0.9	<b>16.4</b>	1.2
	M(1)	M(1)	M(1)	A(1)	A(1)	M(1)	A(1)	A(1)	A(1)	A(1)	M(1)	A(1)
PROMEDIO	10.2	12.5	9.5	0.5	15.2	27.4	0.8	1.3	2.2	0.2	11.0	28.8

PD: pirido[2,3-*d*]pirimidinas, Naph: naftiridin-2(1*H*)-onas, Urea: 7-alkilurea pirido[2,3-*d*]pirimidinas, Indol: indolin-2-onas, Fenil: 1-fenilbenzimidazoles y Anilino: 4-anilinoquinazolininas. Los números en negrita corresponden al peor caso, es decir, el porcentaje de base de datos ordenada necesario para recuperar al menos un representante de cada uno de los *cores*. A y M corresponden a la clasificación de las conformaciones en *active site* y *misdocked*, respectivamente. El número en paréntesis indica si se detecta (1) o no (0) un puente de hidrógeno con los residuos de FGFR implicados en el modo de unión.

Inspeccionando la Tabla 5.7 se muestra claramente que 1fgi y 2fgi difieren en su capacidad para identificar quimiotipos diversos.

Para 1fgi, únicamente se encuentran las indolinonas en posiciones satisfactorias (0.1-0.3%, excepto 3.9% para AUTODOCK), mientras que el resto de compuestos aparecen, en promedio, tras el primer 10% de la base de datos cribada. Asimismo, únicamente el *docking* del representante de este *scaffold* reproduce correctamente el modo de unión con todos los algoritmos/funciones de *scoring*. Lo mismo sucede con los 1-fenilbenzimidazoles, excepto en el caso de los protocolos CHEMSCORE-RESCORE y GOLDS CORE. Para el resto de series químicas, no existe una correlación consistente entre la capacidad para reproducir el modo de unión y su posición en el ranking.

Por una parte, la situación del cribado con CHEMSCORE sería la esperable *a priori* para un algoritmo de *docking*, en el sentido de que los cuatro *scaffolds* para los que no se predice correctamente el modo de unión se recuperan a porcentajes más altos de la lista de compuestos. Sin embargo, en el caso de los cribados con *constraints* de puente de hidrógeno, estos mismos cuatro *misdocked scaffolds*, por el hecho de establecer de algún modo un puente de hidrógeno con los residuos de FGFR requeridos, reciben un elevado *score* que hace que ocupen posiciones preferenciales en la base de datos ordenada. Esto sugiere que el comportamiento de CHEMSCORE-RESTRAINT y GOLDS CORE-RESTRAINT para 1fgi en la recuperación de un *pool* diverso de *scaffolds* es una cuestión de suerte.

Analizando los cribados frente a 2fgi, se identifica, en promedio, un miembro de cuatro de las series activas (pirido[2,3-*d*]pirimidinas, naftiridin-2(1*H*)-onas, 7-alquileura pirido[2,3-*d*]pirimidinas e indolin-2-onas) dentro del primer 2% de la lista de compuestos. Para todas estas conformaciones, se encuentra un modo de unión fiable en todos los casos. Los *scaffolds* en posiciones más retrasadas del ranking (1-fenilbenzimidazoles y 4-anilinoquinazolininas) coinciden con conformaciones fallidas, excepto para la serie de 4-anilinoquinazolininas, para la que el *docking* con AUTODOCK identifica fiablemente el modo de unión del primer activo de la serie identificado. Así, aunque el *docking* ciego con AUTODOCK frente a 2fgi fracasa en la predicción del sitio/modo de unión del compuesto **65** (Tabla 5.4) de la serie 4-anilinoquinazolininas, sí que se reconoce el modo de unión cuando el *binding pocket* es especificado.

De este modo, y tal y como se observa en el apartado de reproducción del modo de unión, el uso de un modelo particular del *binding pocket* desvía los resultados del cribado virtual hacia aquellos compuestos similares a aquellos con los que la proteína ha sido co-cristalizada. En este sentido, 2fgi es una estructura más promiscua que 1fgi, ya que es capaz de acomodar el *docking* de una variedad más amplia de ligandos diversos.

No se aprecia una correlación entre los elevados factores de enriquecimiento y la diversidad de los quimiotipos entre los ligandos recuperados. Por ejemplo, el protocolo CHEMSCORE-RESTRAINT-RESCORE frente a 2fgi es el óptimo desde el punto de vista del criterio de enriquecimiento, aunque fracasa en la reproducción de los modos de unión para dos de los seis *scaffolds*. Contrariamente, AUTODOCK, pese a rendir factores de enriquecimiento inferiores, selecciona representantes de cinco de los seis *cores* dentro del primer 2% de la base de datos ordenada.

Se comprueba si los buenos factores de enriquecimiento son causados por una sobrerrepresentación de aquellos compuestos que se orientan y evalúan bien (es decir, por la diferente representación del número de compuestos de cada quimiotipo). Así, se recalculan los factores de enriquecimiento (sin repetir el *docking*) como el promedio de cinco subconjuntos extraídos de Base\_ACTIV\_1. Estos subconjuntos se construyen seleccionando aleatoriamente 20 compuestos de cada uno de los seis *scaffolds* y 2100 compuestos inactivos. La proporción de inactivos se fija de manera que se mantenga constante la proporción de compuestos activos e inactivos en la base de datos inicial (288/5219 *versus* 120/2220). En la Tabla 5.8 se muestra el promedio de valores de *ef* para los cinco subconjuntos en el primer 2%, 5% y 10% de base de

datos cribada. Comparándose con los resultados de la Tabla 5.6, se observa cómo no existen diferencias significativas.

**Tabla 5.8.** Factores de enriquecimiento (*ef*) obtenidos en el primer 2%, 5% y 10% de base de datos cribada frente a FGFR de los cinco subconjuntos (120/2220) extraídos aleatoriamente de Base\_ACTIV\_1.

	2%		5%		10%	
	1fgi	2fgi	1fgi	2fgi	1fgi	2fgi
CHEMSCORE	3.9	4.3	2.1	3.6	1.4	3.0
CHEMSCORE-RESCORE	2.8	8.2	1.9	5.8	1.5	3.7
CHEMSCORE-RESTRAINT	5.4	5.1	2.7	4.2	1.8	3.6
CHEMSCORE-RESTRAINT-RESCORE	3.4	10.6	2.7	7.4	2.5	4.8
GOLDScore	1.4	7.3	1.0	5.6	1.0	3.5
GOLDScore-RESCORE	3.7	5.4	1.9	4.2	1.7	2.9
GOLDScore-RESTRAINT	2.5	10.1	2.3	6.9	1.6	4.4
GOLDScore-RESTRAINT-RESCORE	5.4	5.8	2.8	4.4	1.8	3.6
AUTODOCK	0.5	3.9	1.6	3.2	1.7	3.0

En resumen, desde el punto de vista del enriquecimiento de activos, la combinación de la función CHEMSCORE en el *docking* con *constraints* y el posterior *rescoring* con GOLDScore parece ser la mejor opción en este estudio frente a 2fgi. Sin embargo, AUTODOCK supera a este protocolo en la identificación de quimiotipos diversos.

En un intento de aprovechar la información estructural contenida en las dos estructuras de FGFR (1fgi y 2fgi), se escoge el protocolo AUTODOCK para combinar los resultados del cribado virtual para 1fgi y 2fgi en un esquema de modelo por consenso (*consensus model*). Se adopta la estrategia *rank-by-rank*<sup>392</sup>, según la cual los compuestos candidatos de la base de datos se ordenan por el rango o posición en el ranking promedio previsto por todas las funciones de *scoring*. En este caso, primero se calcula simplemente el rango final de cada uno de los compuestos de Base\_ACTIV\_1 como el rango promedio recibido por los dos cribados con AUTODOCK frente a 1fgi y 2fgi. Los factores de enriquecimiento se incrementan desde 4.0 (2fgi) a 6.6 (*consensus scoring*) en el primer 2% de base de datos. Posteriormente, se repite el modelo por consenso considerando únicamente el rango recibido en aquel cribado que predice correctamente el modo de unión para cada una de las series. Es decir, a las pirido[2,3-*d*]pirimidinas, naftiridin-2(1*H*)-onas, 7-alkilurea pirido[2,3-*d*]pirimidinas y 4-anilinoquinazolininas se les asigna el rango encontrado en el cribado virtual frente a 2fgi, mientras que los 1-fenilbenzimidazoles reciben el rango por su posición en el ranking del cribado frente a 1fgi. Para los compuestos inactivos y las indolin-2-onas, se calcula un rango promedio de los dos cribados. Este segundo modelo rinde valores de *ef* de 12.1, 6.9 y 5.1 en el primer 2%, 5% y 10% de la base de datos muestreada, respectivamente. Además, se encuentra al menos un compuesto representativo de cada quimiotipo dentro del primer 2% de base de datos, con la excepción de los 1-fenilbenzimidazoles. Así, el *consensus scoring* permite incrementar el mejor enriquecimiento encontrado con una única estructura (2fgi con CHEMSCORE-RESTRAINT-RESCORE).



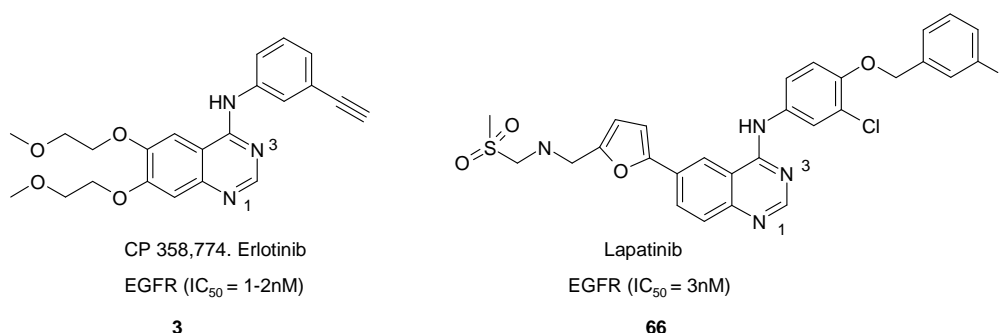
## 5.2. *Docking* frente a EGFR

En este apartado se analizan los citados tres puntos (predicción del modo de unión, predicción de sitio de unión y cribado virtual) frente a EGFR.

### 5.2.1. Preparación de las estructuras cristalinas de EGFR

En el momento de la realización del trabajo, se encuentran depositadas en el *Protein Data Bank* tres estructuras del dominio tirosina quinasa de EGFR (ErbB-1 o HER1):

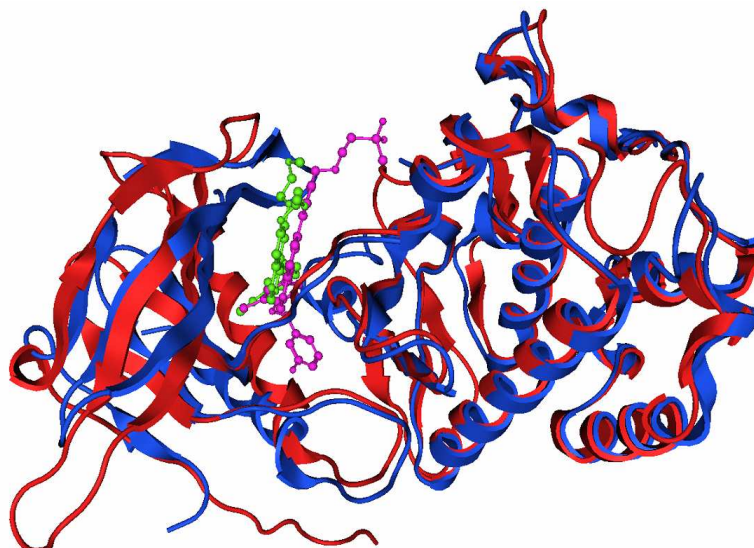
- La forma apo de EGFR, entrada PDB 1m14, con resolución 2.60 Å.<sup>401</sup> Se encuentra en una conformación activa.
- Complejado con erlotinib (OSI-774, CP-358,774, Tarceva<sup>TM</sup>), una 4-anilinoquinazolina (**3**) aprobada actualmente por la FDA para el tratamiento de cáncer de pulmón no microcítico (CPNM). Entrada PDB 1m17, con resolución 2.60 Å.<sup>401</sup>
- Co-cristalizado con lapatinib (GW572016), una 4-anilinoquinazolina (**66**) en desarrollo clínico. Entrada PDB 1xkk, con resolución 2.40 Å.<sup>402</sup>



**Figura 5.12.** Estructuras de los inhibidores erlotinib y lapatinib para los que existen estructuras resueltas co-cristalizadas con EGFR.

Recientemente (junio 2006), se han depositado también otras tres estructuras de EGFR activo, libre y complejado con un conjugado peptídico análogo al ATP, y una estructura inactiva en complejo con AMP-PNP (adenililimidodifosfato).<sup>403</sup>

Erlotinib y lapatinib se unen a EGFR de manera diferente.<sup>402</sup> Aunque el plegamiento general de la estructura de EGFR en 1xkk, 1m14 y 1m17 es muy similar, existen diferencias significativas en la orientación de los dos lóbulos N-terminal y C-terminal. La hendidura de unión del ATP en lapatinib/EGFR está en una conformación relativamente cerrada, análoga a una conformación inactiva. En el complejo erlotinib/EGFR, la conformación está mucho más abierta. Respecto al *loop* de activación (residuos 831-860), apo-EGFR y erlotinib/EGFR tienen conformaciones similares a las de las quinazinas activas, mientras que lapatinib/EGFR presenta una conformación más próxima al estado inactivo. En la Figura 5.13 se muestra la superposición de las estructuras 1m17 (*rojo*) y 1xkk (*azul*), con una RMSD de 10.4 Å. Para el estudio de *docking*, se escogen las estructuras 1m17 y 1xkk.

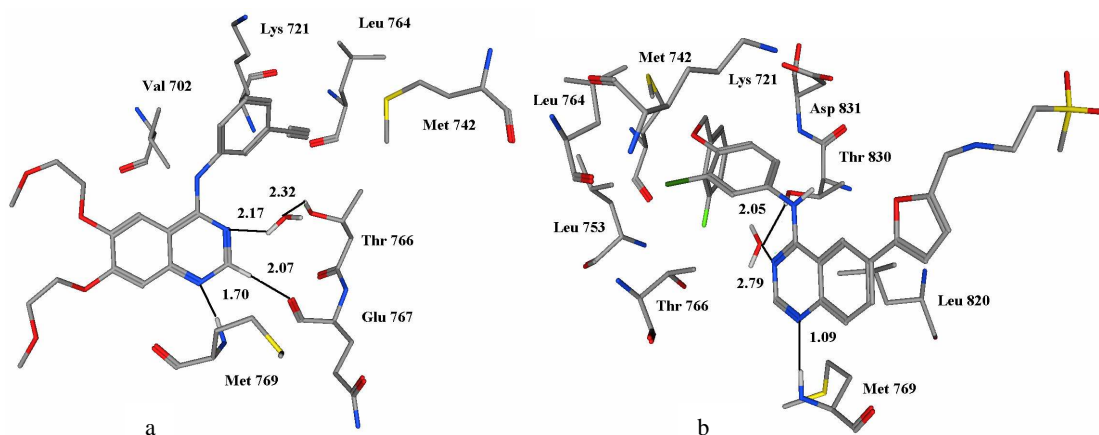


**Figura 5.13.** Superposición de la estructura 1m17 (rojo) y 1xkk (azul) de EGFR. Se muestran sus respectivos inhibidores co-cristalizados: erlotinib (verde) y lapatinib (rosa).

También los anillos quinazolina de los dos compuestos interactúan por puente de hidrógeno con EGFR de manera diferente. En ambos inhibidores, el nitrógeno N1 de la quinazolina (Figura 5.12) acepta un puente de hidrógeno del grupo NH de la cadena principal del residuo Met<sup>769</sup>. El nitrógeno N3 de erlotinib establece un puente de hidrógeno mediado por una molécula de agua con la cadena lateral del aminoácido Thr<sup>766</sup>. Lapatinib establece este puente de hidrógeno de su N3, también mediado por agua, pero con Thr<sup>830</sup>. Desde un punto de vista mecanístico, se ha propuesto que estas diferencias estructurales son las responsables de las diferentes tasas de disociación encontradas en ambos inhibidores reversibles.<sup>402</sup>

Anteriormente a la resolución de estas estructuras, varios autores habían propuesto el modo de interacción de las 4-anilinoquinazolininas con EGFR a partir de modelos construidos por homología, proponiéndose tanto la interacción mediada por agua por puente de hidrógeno con Thr<sup>830</sup> (Wissner<sup>404</sup>) como con la Thr<sup>766</sup> (Hou<sup>405</sup>).

En la Figura 5.14 se muestran las interacciones de erlotinib con la estructura 1m17 (a) y de lapatinib con 1xkk (b), respectivamente.



**Figura 5.14.** (a) Interacción de erlotinib en el sitio activo de 1m17. Interacciones por puente de hidrógeno con el residuo Met<sup>769</sup> (distancia 2.70 Å) y una molécula de agua (distancia 2.78 Å). (b) Interacción de lapatinib en el sitio activo de 1xkk. Interacciones por puente de hidrógeno con el residuo Met<sup>769</sup> (distancia 4.92 Å) y una molécula de agua (distancia 2.79 Å). El nitrógeno de la cadena lateral también establece una interacción por puente de hidrógeno con una molécula de agua.

Como en el caso de FGFR, para los dos complejos (1m17 y 1xkk) se añaden todos los hidrógenos, relajándose la posición de los mismos siguiendo el procedimiento descrito en el apartado 5.1.1. Ambas estructuras contienen una única cadena A.

Para esta diana, la inclusión de las aguas estructurales importantes en la interacción resulta imprescindible en el *docking*, tal y como han apuntado Aparna y colaboradores<sup>406</sup> en un estudio de cribado virtual de 4-anilinoquinazolininas frente a EGFR (1m17) con GOLD.

- Para la estructura 1m17, se mantiene únicamente el residuo de agua estructural (water 10 en el fichero *pdb*) implicado en la interacción con Thr<sup>766</sup>, en lugar de las cinco aguas estructurales presentes en la estructura y que no se encuentran próximas al ligando.
- Para 1xkk, se eliminan 74 de las 77 moléculas de agua presentes, excepto tres residuos posicionados en el sitio de unión (water 5, water 4 y water 22 en el fichero *pdb*), uno de los cuales establece la interacción esencial con la Thr<sup>830</sup>.

Para cada complejo, se optimizan las posiciones iniciales de sus ligandos nativos y se generan los *cross-decoys* y las posiciones iniciales del *docking* ciego según el procedimiento descrito en 5.1.1.

### 5.2.2. Resultados de la predicción del modo de unión en EGFR: estructuras nativas y *cross-decoys*

Esta predicción se realiza con AUTODOCK y GOLD en las condiciones especificadas en el apartado 5.1.2. Únicamente se modifica:

- **AUTODOCK.** Se disminuye el tamaño de la malla con el fin de reducir el tiempo de cálculo. Se prueban tamaños de malla de 61×61×61 y 101×101×101. Los ligandos se parametrizan con las cargas Gasteiger-Marsili, calculándose los potenciales de afinidad atómica estándar y el conjunto adicional (halógenos, nitrógeno aceptor de puente de hidrógeno...). Para modelar las aguas éstas se funden en la estructura proteica, calculándose los mapas de afinidad para la combinación proteína-aguas, siguiendo el proceso aconsejado en la referencia [407].
- **GOLD.** Por disponibilidad del programa, a partir de este punto se utiliza la versión 3.0.1. En ella, se ha introducido de un nuevo término de *scoring* para evaluar las interacciones débiles por puente de hidrógeno C-H...O. Éstas son muy comunes en muchos complejos de quinasas con sus inhibidores<sup>216</sup>, como es el caso del carbono C-2 de erlotinib con el oxígeno del carbonilo de Glu<sup>767</sup> (Figura 5.14.a).

En la Tabla 5.9 se muestra la RMSD de superposición de la primera conformación en el ranking respecto a la conformación cristalográfica del ligando (promedio y desviación estándar de cinco repeticiones) para las estructuras nativas y los *cross-decoys*, según los cuatro protocolos. Como en la Tabla 5.1, la intensidad del sombreado indica la calidad de la conformación obtenida (gris oscuro: RMSD < 2 Å, gris claro: 2-4 Å) y la negrita, las estructuras nativas.

**Tabla 5.9.** Resultados de RMSD (Å) y sus desviaciones estándar (paréntesis) para las posiciones inicialmente situadas en el sitio de unión en el *docking* y el *cross-docking* frente a las dos estructuras resueltas de EGFR para los cuatro protocolos.

		Erlotinib	Lapatinib
AUTODOCK	1m17	<b>1.7 (0.1)</b>	4.2 (0.5)
Mapas de afinidad extra. Gasteiger-Marsili. 61×61×61	1xkk	1.8 (0.1)	<b>2.1 (0.2)</b>
AUTODOCK	1m17	<b>1.7 (0.1)</b>	4.4 (0.6)
Mapas de afinidad extra. Gasteiger-Marsili. 101×101×101	1xkk	1.8 (0.2)	<b>1.7 (0.5)</b>
GOLD-GOLDScore	1m17	<b>5.2 (3.7)</b>	9.8 (4.0)
	1xkk	3.8 (2.7)	<b>7.6 (0.3)</b>
GOLD-CHEMSCORE	1m17	<b>3.1 (1.9)</b>	7.5 (0.6)
	1xkk	4.9 (3.6)	<b>2.5 (0.8)</b>

AUTODOCK y CHEMSCORE encuentran la conformación activa tanto para erlotinib en 1m17 como para lapatinib en 1xkk. Se destaca que en el *docking* de erlotinib frente a 1m17 con CHEMSCORE, el promedio de 3.1 se obtiene porque una de las cinco repeticiones se encuentra a 6.6 Å de la conformación cristalográfica, superponiéndose las restantes cuatro soluciones con una RMSD inferior a 2.2 Å. GOLDScore fracasa claramente en la identificación del modo de unión para lapatinib en 1xkk (mala orientación de su cadena lateral), comportándose ligeramente mejor para erlotinib en 1m17. De hecho, el valor promedio de RMSD de 5.2 Å se debe a que tres de las cinco soluciones se sitúan por encima de los 4 Å, reproduciéndose adecuadamente el modo de unión (< 2 Å) únicamente en dos de ellas.

En el estudio de Aparna<sup>406</sup>, los autores presentan valores de RMSD inferiores, aunque debe considerarse que únicamente incluyen el esqueleto quinazolina en el cálculo de la RMSD. Tampoco se especifica si se selecciona exclusivamente la primera conformación del ranking o cuál de las dos funciones de GOLD es la que proporciona estos resultados. En cualquier caso, en un intento de reproducir sus resultados óptimos para GOLDScore, se repite el *docking* de erlotinib en 1m17 empleando la definición de sitio activo utilizada por estos autores, que incluye un listado de residuos particular. Los resultados obtenidos no varían ostensiblemente respecto a los mostrados en la Tabla 5.9.

Respecto a los resultados de los *cross-decoys*, únicamente erlotinib, con sustituyentes de menor tamaño en la cadena lateral que lapatinib, es capaz de reproducir su modo de unión en 1xkk con AUTODOCK. Con las funciones CHEMSCORE y GOLDScore, en tres de las cinco repeticiones se obtienen conformaciones con una RMSD comprendida entre 2.3 y 2.5 Å.

### 5.2.3. Resultados del *docking* ciego en EGFR

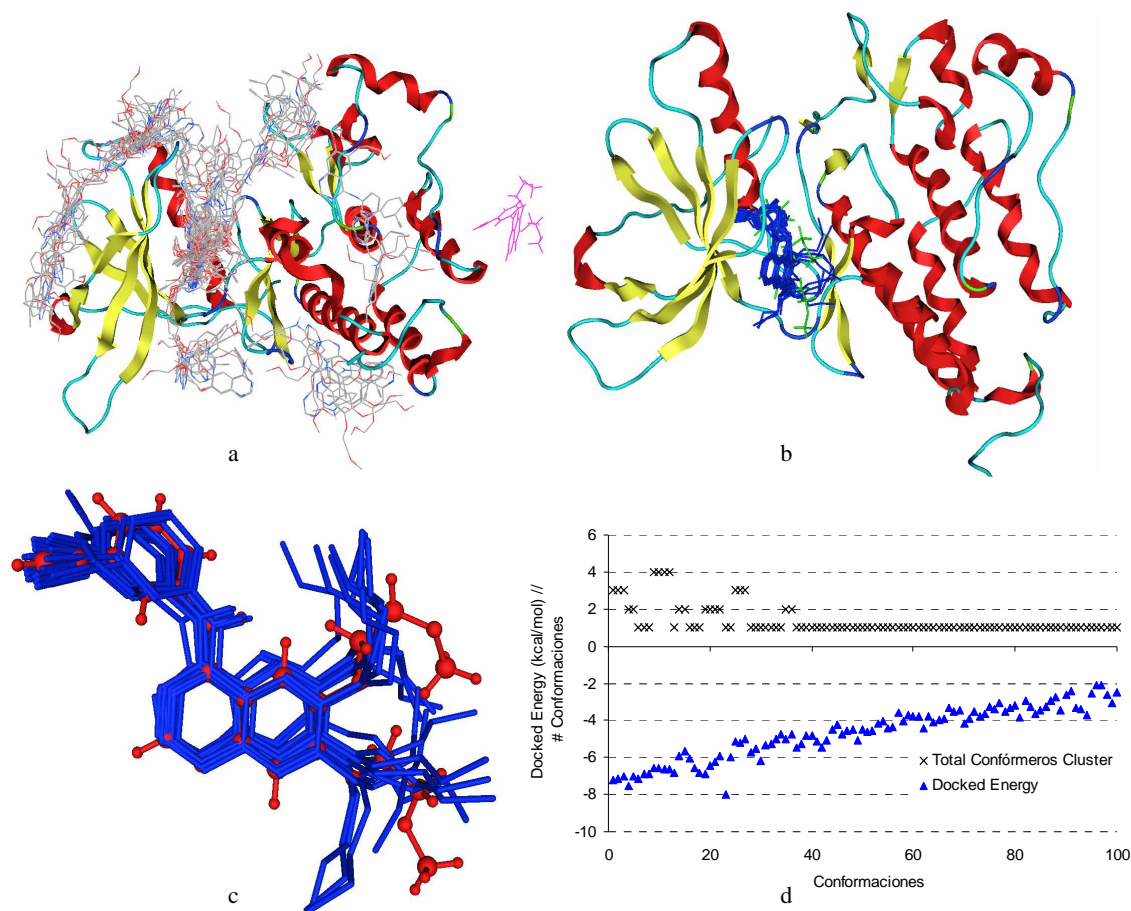
Se realiza el *docking* ciego de erlotinib y lapatinib en sus estructuras nativas y los correspondientes *cross-decoys* con AUTODOCK, posicionando los ligandos en la superficie de la proteína y calculando los mapas de afinidad para un tamaño de malla de 181×181×181.

En la Tabla 5.10 se muestra la RMSD de superposición de la primera conformación del ranking de las 100 conformaciones obtenidas con AUTODOCK en cinco *dockings* ciegos respecto de la estructura cristalina.

**Tabla 5.10.** Resultados de RMSD (Å) y sus desviaciones estándar (paréntesis) para el *docking* ciego y el *cross-docking* ciego de las dos estructuras de EGFR con AUTODOCK.

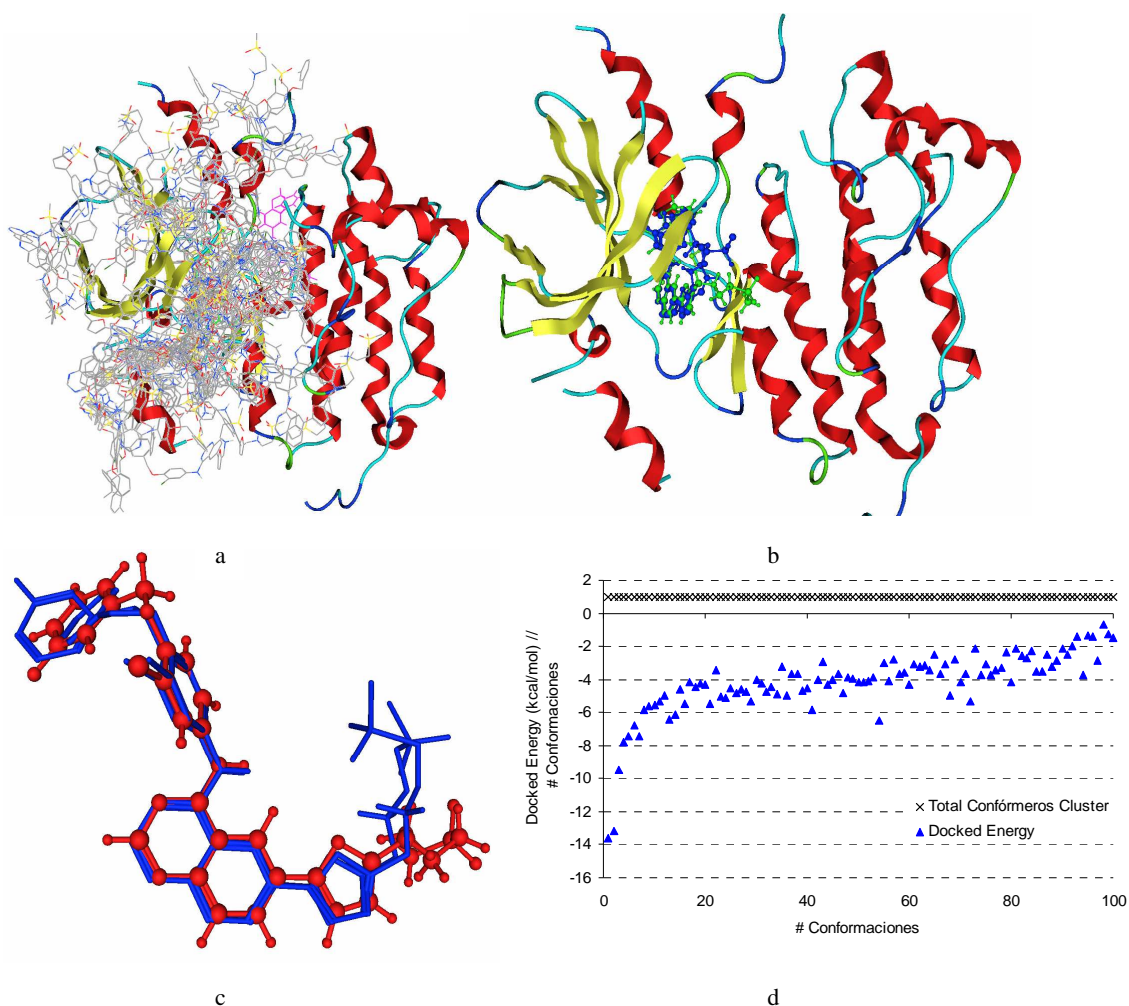
		Erlotinib	Lapatinib
AUTODOCK	1m17	1.7 (0.1)	5.5 (2.9)
Mapas de afinidad extra. Gasteiger-Marsili. 181×181×181	1xkk	1.7 (0.1)	4.0 (3.1)

El *docking* ciego identifica correctamente el sitio y el modo de unión para erlotinib en 1m17. En la Figura 5.15.a se muestran los 100 conforméromos de un *docking*. De ellos, los 12 primeros del ranking (*azul*) se superponen sobre la estructura co-cristalizada (*rojo*) (Figura 5.15.c). Éstos pertenecen a los *clusters* más poblados (Figura 5.15.d).



**Figura 5.15.** Resultados del *docking* ciego para erlotinib frente a 1m17 con AUTODOCK y el protocolo con cargas Gasteiger-Marsili. (a) 100 conformaciones resultantes. (b) 12 conformaciones de menor *docked energy* que detectan el sitio activo. (c) Superposición de las primeras conformaciones del ranking (*azul*) frente a erlotinib co-cristalizado (*rojo*). (d) Distribución de las 100 conformaciones en los *clusters* y su energía de *docking* respectiva (*docked energy*). Los 100 conforméromos se ordenan en función de su *binding energy*.

Por otra parte, también el *cross-docking* ciego de erlotinib en 1xkk es exitoso. Para la proteína nativa de lapatinib, 1xkk, en dos casos de cinco repeticiones, la primera conformación se encuentra a una RMSD de superposición de la estructura cristalográfica inferior a 2 Å. La Figura 5.16 muestra los resultados para uno de estos dos casos. Los dos primeros de los 100 conforméromos se superponen frente a la estructura co-cristalizada con una RMSD inferior a 2.0 Å (Figura 5.16.c). Éstos presentan un valor absoluto de energía de *docking* claramente superior al resto de conformaciones (Figura 5.16.d).



**Figura 5.16.** Resultados del *docking* ciego para lapatinib frente a 1xkk con AUTODOCK y el protocolo con cargas Gasteiger-Marsili. (a) 100 conformaciones resultantes. (b) 2 conformaciones de menor *docked energy* que encuentran el sitio activo. (c) Superposición de las dos primeras conformaciones del ranking (*azul*) frente a lapatinib co-cristalizado (*rojo*). (d) distribución de las 100 conformaciones en los *clusters* y su energía de *docking* respectiva (*docked energy*).

#### 5.2.4. Cribado virtual en EGFR

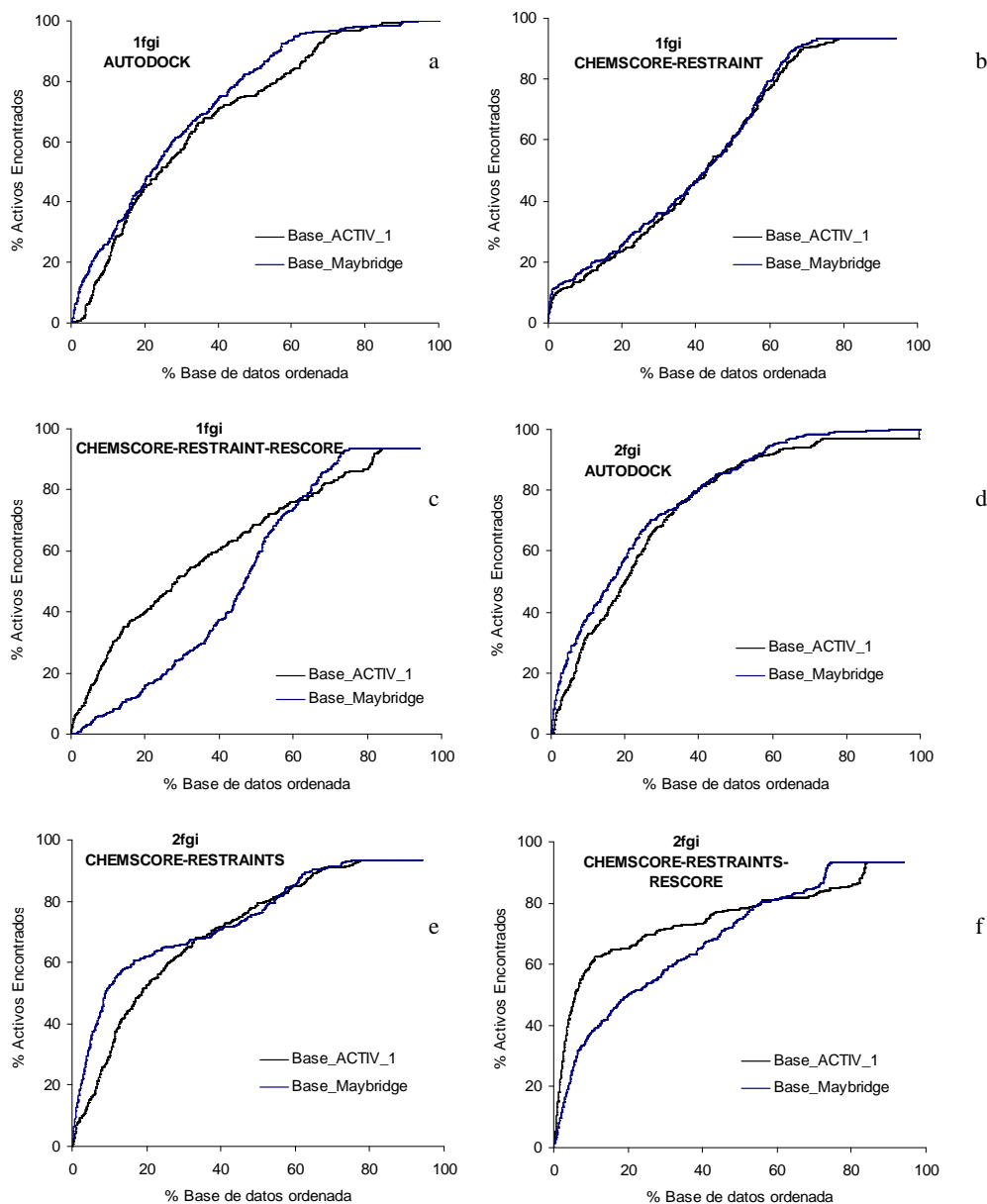
La base de datos COBRA se utilizó durante una estancia realizada en el grupo del profesor Schneider en la Johann Wolfgang Goethe-Universität en Frankfurt. Dado que el trabajo que sigue se realiza en el IQS, se utiliza desde este punto otro *pool* de 4342 compuestos inactivos extraídos del catálogo Maybridge. Aunque la selección de los compuestos es aleatoria, se construye el nuevo *pool* de manera que corresponda a una quimioteca focalizada, en el sentido de que presente propiedades 1D similares a las del *pool* de 288 activos (Tabla 5.11). En adelante, se denomina al conjunto total de activos e inactivos Base\_Maybridge.

**Tabla 5.11.** Propiedades 1D del *pool* de activos e inactivos de la base de datos Base\_Maybridge.

	Peso Molecular	b_1rotN	a_acc	a_don
288 Activos	416.0 (146.4)	8.5 (4.0)	3.6 (1.3)	1.9 (1.3)
4342 Inactivos	415.0 (40.1)	6.9 (1.3)	3.5 (1.4)	1.5 (0.7)



Antes de proceder al cribado virtual frente a EGFR, se repite el cribado virtual de Base\_Maybridge frente a FGFR para los dos protocolos óptimos encontrados en el cribado de Base\_ACTIV\_1 (Figura 5.11): AUTODOCK y CHEMSCORE-RESTRAINTS (con y sin *rescoring*). El objetivo es constatar la posible influencia del *pool* de inactivos en el cribado. Las curvas de enriquecimiento de Base\_Maybridge, comparadas con las de Base\_ACTIV\_1, se muestran en la Figura 5.17 para 1fgi y 2fgi.



**Figura 5.17.** Curvas de enriquecimiento del cribado virtual frente a FGFR (1fgi y 2fgi) de las bases de datos Base\_ACTIV\_1 y Base\_Maybridge. (a) AUTODOCK-1fgi. (b) CHEMSCORE-RESTRAINT-1fgi. (c) CHEMSCORE-RESTRAINT-RESCORE-1fgi. (d) AUTODOCK-2fgi. (e) CHEMSCORE-RESTRAINTS-2fgi. (f) CHEMSCORE-RESTRAINTS-RESCORE-2fgi.

Se aprecia cómo no existen apenas diferencias para los cribados con las funciones AUTODOCK y CHEMSCORE-RESTRAINT, indicando que la presunción de compuestos inactivos en ambas bases de datos no desvía significativamente el resultado. Sin embargo, existen notables diferencias al aplicar la función de *rescoring*, en los que Base\_Maybridge se sitúa muy por debajo de Base\_ACTIV\_1. Dado que los resultados sin *rescoring* son muy similares para las dos bases de datos, parece que el motivo de esta variación sea la propia función de *rescoring*.

En este sentido, se enfatiza que Base\_ACTIV\_1 está cribada con GOLD 2.0, mientras que Base\_Maybridge lo está con GOLD 3.0.1. Dada la diferente disponibilidad de base de datos/versión del programa, no se puede repetir el cálculo de Base\_ACTIV\_1 con GOLD 3.0.1 para comprobar si las diferencias se deben al programa o a la base de datos.

El cribado virtual frente a EGFR de Base\_Maybridge se realiza con los dos programas AUTODOCK y GOLD para las dos estructuras 1m17 y 1xkk. Del *pool* de activos ACTIV\_1, compuesto por 288 inhibidores de FGFR, se extraen aquellos compuestos cuya actividad ha sido determinada frente a EGFR y que presentan una IC<sub>50</sub> inferior a 50 µM. En total, se contabilizan 69 compuestos inhibidores de EGFR: las 29 4-anilinoquinazolinonas, 10 7-alquilurea pirido[2,3-*d*]pirimidinas y 30 pirido[2,3-*d*]pirimidinas. Sus estructuras y valores de IC<sub>50</sub> frente a EGFR se muestran en el Anexo 2.

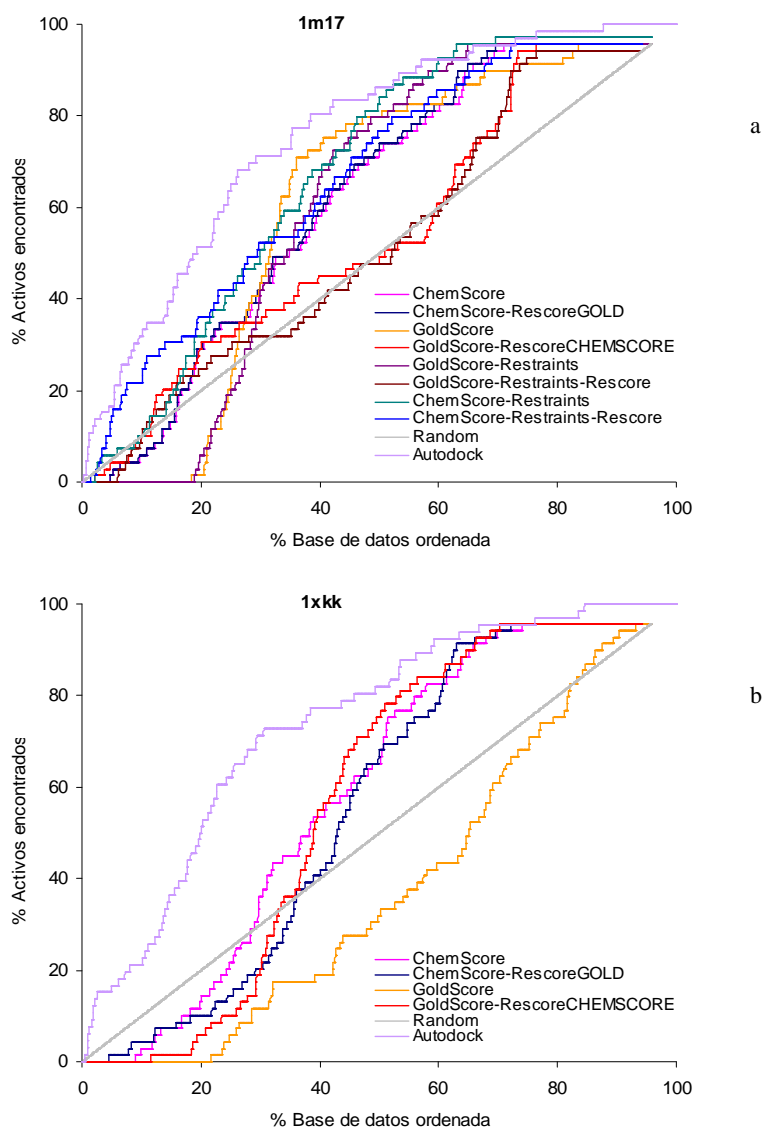
Así, la base de datos utilizada consiste de 69 activos y 4342 inactivos (4411 compuestos), presentando un factor de enriquecimiento máximo de 63.9 (4411/69). En adelante, se denomina a esta base de datos Base\_Maybridge\_69.

Los protocolos estudiados y sus condiciones son:

- **AUTODOCK.** Cargas Gasteiger-Marsili y con mapas de afinidad adicionales para los tipos atómicos. Tamaño de malla de 101×101×101, se requieren 10 confórmers por compuesto.
- **GOLD.** Para los cuatro protocolos descritos en el apartado 5.1.5 (GOLDScore, CHEMScore, GOLDScore-REScore, CHEMScore-REScore), se analiza la influencia de la presencia o ausencia de restricciones de tipo “*Protein Hydrogen Bonds*”, con un factor de peso de 10. GOLD únicamente permite fijar átomos de la proteína con carácter aceptor. Por ello, la interacción por puente de hidrógeno con el nitrógeno de la cadena principal de Met<sup>769</sup> como dador, común a 1m17 y 1xkk, no puede introducirse como restricción. Tampoco las moléculas de agua mediadoras de puente de hidrógeno. Para 1m17 se requiere que el residuo Gln<sup>767</sup> actúe como aceptor, ya que éste establece un contacto por puente de hidrógeno C-H...O con erlotinib (Figura 5.14.a). Para 1xkk, se analiza si Gln<sup>767</sup> interviene en un enlace débil tipo C-H...O como en 1m17, pero no se detecta. De todos modos, el oxígeno carbonílico de este residuo no se encuentra accesible al solvente en 1xkk, por lo que no se puede fijar como restricción. Así, para 1xkk no existe ningún residuo de la proteína que actúe como dador, por lo que no se aplican restricciones con esta estructura. También se destaca el hecho de que al aplicar un *rescoring*, GOLD no considera las aguas.

En resumen, en total se examinan un total de nueve protocolos diferentes para 1m17 y cinco para 1xkk. En la Figura 5.18 se muestran las curvas de enriquecimiento obtenidas en el *docking* frente a 1m17 (a) y 1xkk (b).





**Figura 5.18.** Curvas de enriquecimiento obtenidas con los protocolos de *docking* frente EGFR para las estructuras (a) 1m17 y (b) 1xkk en el cribado de Base\_Maybridge\_69. La línea gris representa los valores obtenidos si los activos se seleccionasen aleatoriamente.

En la Tabla 5.12 se muestran los factores de enriquecimiento (*ef*) en el primer 2%, 5% y 10% de base de datos Base\_Maybridge\_69 cribada.

**Tabla 5.12.** Factores de enriquecimiento (*ef*) obtenidos en el primer 2%, 5% y 10% de base de datos Base\_Maybridge\_69 cribada frente EGFR.

	2%		5%		10%	
	1m17	1xkk	1m17	1xkk	1m17	1xkk
CHEMSCORE	0.0	0.0	0.3	0.0	0.6	0.3
CHEMSCORE-RESCORE	0.0	0.0	0.3	0.3	0.6	0.4
CHEMSCORE-RESTRAINT	0.0		1.2		1.0	
CHEMSCORE-RESTRAINT-RESCORE	0.7		2.9		2.2	
GOLDSCORE	0.0	0.0	0.0	0.0	0.0	0.0
GOLDSCORE-RESCORE	0.0	0.0	0.9	0.0	0.9	0.0
GOLDSCORE-RESTRAINT	0.0		0.0		0.0	
GOLDSCORE-RESTRAINT-RESCORE	0.0		0.0		1.0	
AUTODOCK	6.0	6.0	3.3	3.3	3.2	2.1

Con excepción de AUTODOCK, todos los protocolos de GOLD funcionan realmente mal en la discriminación de compuestos activos/inactivos, exceptuando ligeramente el protocolo CHEMSCORE-RESTRAINT-RESCORE frente a 1m17. Desde un punto de vista de enriquecimiento, con los resultados de AUTODOCK apenas existen diferencias entre las estructuras 1m17 y 1xkk.

En la Tabla 5.13 se muestran los porcentajes de base de datos muestreada a los que se encuentra el primer representante de cada uno de los tres quimiotipos contenidos en el *pool* de activos, así como si la orientación de estos compuestos reproduce (A) o no (M) el modo de unión previsible. Como en el caso de FGFR, se presenta la existencia (1) o no (0) de un puente de hidrógeno entre el correspondiente compuesto y los residuos clave de EGFR (principalmente Met<sup>769</sup>)

**Tabla 5.13.** Porcentaje de base de datos cribada frente a EGFR para la que al menos un representante activo de cada uno de los tres *scaffolds* de Base\_Maybridge\_69 es recuperado.

	1m17			1xkk		
	PD	Urea	Anil	PD	Urea	Anil
CHEMSCORE	4.7 M(0)	<b>49.9</b> <b>M(0)</b>	10.9 M(0)	11.8 M(1)	28.3 M(0)	<b>8.9</b> <b>A(1)</b>
CHEMSCORE-RESCORE	4.7 M(0)	<b>49.2</b> <b>M(0)</b>	10.8 M(0)	7.9 M(1)	<b>33.8</b> <b>M(0)</b>	4.4 A(1)
CHEMSCORE-RESTRAINT	2.1 M(1)	9.5 M(1)	<b>17.5</b> <b>M(1)</b>			
CHEMSCORE-RESTRAINT-RESCORE	1.4 M(1)	<b>10.1</b> <b>M(1)</b>	<b>10.1</b> A(1)			
GOLDSCORE	9.3 M(1)	18.4 M(0)	<b>21.0</b> <b>M(1)</b>	28.6 M(0)	63.2 M(0)	<b>64.0</b> <b>M(0)</b>
GOLDSCORE-RESCORE	2.1 M(1)	<b>20.0</b> <b>M(1)</b>	11.6 M(0)	18.6 M(1)	<b>41.6</b> <b>M(0)</b>	11.6 M(0)
GOLDSCORE-RESTRAINT	19.1 M(0)	<b>19.6</b> <b>M(0)</b>	18.9 M(0)			
GOLDSCORE-RESTRAINT-RESCORE	6.0 A(1)	<b>21.7</b> <b>A(1)</b>	20.9 M(0)			
AUTODOCK	3.2 A(1)	0.7 A(1)	<b>5.5</b> <b>A(1)</b>	<b>4.9</b> <b>A(1)</b>	0.5 A(1)	1.8 A(1)
PROMEDIO	5.8	22.1	14.1	14.4	33.5	22.3

PD: pirido[2,3-*d*]pirimidinas, Urea: 7-alkilurea pirido[2,3-*d*]pirimidinas, Anilino: 4-anilinoquinazolininas. Los números en negrita corresponden al peor caso, es decir, el porcentaje de base de datos ordenada necesario para recuperar al menos un representante de cada uno de los *cores*. A y M corresponden a la clasificación de las conformaciones en *active site* y *misdocked*, respectivamente. El número en paréntesis indica si se detecta (1) o no (0) un puente de hidrógeno con los residuos de EGFR implicados en el modo de unión.

Básicamente, no se aprecian diferencias contundentes entre las dos estructuras de EGFR. Los protocolos con GOLD apenas identifican alguna conformación activa, existiendo un acuerdo entre la falta de reproducción del modo de unión y los bajos enriquecimientos obtenidos.

AUTODOCK identifica la conformación correcta para al menos un representante de cada uno de los tres quimiotipos, tanto en el *docking* frente a 1m17 como frente a 1xkk.

Se combinan los dos cribados virtuales con AUTODOCK frente a 1m17 y 1xkk en un esquema por consenso siguiendo la estrategia *rank-by-rank*. Los factores de enriquecimiento obtenidos en el primer 2%, 5% y 10% de la base de datos muestreada son 6.8, 3.3 y 2.7, respectivamente. Comparando con los mostrados en la Tabla 5.12, la estrategia *rank-by-rank* supone únicamente un leve enriquecimiento (6.0 *versus* 6.8).

### 5.3. *Docking* frente a PDGFR

En este apartado se analizan los citados tres puntos (predicción del modo de unión, predicción de sitio de unión y cribado virtual) frente a PDGFR. Dado que no existe ninguna estructura cristalina resuelta del dominio tirosina quinasa de dicha PDGFR, se modela por homología.

#### 5.3.1. Modelización por homología del dominio tirosina quinasa de PDGFR

El PDGFR  $\alpha$  y  $\beta$  es un miembro de la subfamilia tipo III de los RTKs transmembrana, a la que también pertenecen el receptor FMS (*colony-stimulating factor-1 receptor*, CSF-1R), el receptor Flt-3 y el receptor c-Kit. De los dos receptores de la familia PDGFR (Tabla 2.1), se ha considerado la secuencia de PDGFR- $\beta$  humana (entrada SWISS-PROT P09619 y Protein Kinase Resource\_ID, PKR\_ID: 1007936). Esta secuencia consta de 1106 aminoácidos, distribuidos en:

- 1-32: péptido señal
- 33-531: dominio extracelular
- 532-556: dominio transmembrana
- 557-1106: dominio citoplasmático. El dominio tirosina quinasa comprende 363 residuos: desde el aminoácido 600 hasta el 962. Está dividido por el *kinase insert domain* (KID) que consta de 104 residuos (desde 695-798).<sup>408,409</sup>. Las regiones significativas se encuentran en:
  - *Loop* de unión del nucleótido: consistente de los aminoácidos 606-614 (secuencia Gly-X-Gly-XX-Gly)
  - Residuo Lys<sup>634</sup> implicado en la unión de ATP
  - Residuo Asp<sup>826</sup>: sitio activo
  - Fosfotirosinas Tyr<sup>751</sup> y Tyr<sup>857</sup> (autocatálisis)

Se han publicado diversos modelos de PDGFR. Inicialmente, se basaron en la estructura de la PKA (entrada PDB 1atp), como los descritos por el grupo de investigación de Parke-Davis Pharmaceutical.<sup>346,358</sup> Diller<sup>381</sup> y colaboradores construyeron el modelo de PDGFR- $\beta$  a partir de las estructuras de FGFR (entrada PDB 1agw), VEGFR-2 (entrada PDB 1vr2) y IGFR (entrada PDB 1ir3). Böhmer *et al*<sup>410</sup> modelaron el PDGFR- $\beta$  a partir de las estructuras de VEGFR-2 (entrada PDB 1vr2), FGFR-1 (entradas PDB 1agw, 1fgi, 1fgk, 2fgi) y c-Abl quinasa (entradas PDB 1fpu y 1iep) e IGFR (entrada PDB 1irk).

La versión de MOE 2004 contiene una base de datos de dominios quinasa diseñada para realizar las búsquedas de plantillas en modelización. Su uso se desestima ya que, por una parte no se encuentra perfectamente actualizada y, además, se observan algunas diferencias de secuencia importantes en diversas entradas respecto a las de la entrada PDB de las que han sido extraídas.

Se opta por realizar una búsqueda de las plantillas estructurales en el *Protein Data Bank* con el servidor FASTA (apartado 1.5.1) utilizando la secuencia del dominio quinasa de PDGFR- $\beta$ . Se impone como valor de corte un *fasta score* inferior a  $1E-21$ . Se comparan los resultados con los devueltos por el servidor *SWISS-MODEL Blast*<sup>133</sup>, que trabaja con la base de datos ExNRL-3D (derivada del PDB). Ambas búsquedas son muy similares, difiriendo únicamente en el orden relativo de aparición de ciertas secuencias. En la Tabla 5.14 se listan las plantillas, siguiendo el orden propuesto por FASTA, considerándose únicamente las cadenas A de cada una de las estructuras.

**Tabla 5.14.** Plantillas propuestas por FASTA para la secuencia del dominio quinasa de PDGFR- $\beta$ .

<b>Entrada PDB</b>	<b>Tirosina Quinasa</b>	<b>Res (Å)</b>	<b>Residuos Incluidos Cadena A</b>	<b>Origen Fuente</b>	<b>Descripción conformación. Ligandos</b>
1t45 <sup>411</sup>	c-Kit	1.90	547-935 (694-753: KID)	Humano	Autoinhibido. Sin fosforilar
1t46 <sup>411</sup>	c-Kit	1.60	565-933 (748-761) (933-934) (694-753: KID)	Humano	Unido a STI-571. Conformación inactiva
1pkg <sup>412</sup>	c-Kit	2.90	567-927 (752-762)	Humano	Unido a ADP-MG. Conformación activa.
1rjb <sup>413</sup>	Flt-3	2.10	572-947 (649-654) (762-782)	Humano	Autoinhibido. Sin fosforilar
1vr2 <sup>414</sup>	VEGFR-2 (Flk1)	2.40	820-1168 (806-819) (990-998) (1048-1083) (1169-1171) (940-989)	Humano	Sin fosforilar. Conformación inactiva. Sin ligando.
1gjo	FGFR2 (bek)	2.40	465-768 (453-466) (504-506) (581-595)(765-768)	Humano	No publicado
1fgi <sup>375</sup>	FGFR1 (Flg)	2.50	464-762 (501-504) (580-591)	Humano	Unido a SU5402
1agw <sup>375</sup>	FGFR1 (Flg)	2.40	464-762 (486-490)(501-504) (580-591)	Humano	Unido a SU4984
1fgk <sup>415</sup>	FGFR1 (Flg)	2.0	464-762 (456-463) (486-490) (501-504) (580-591) (763-765)	Humano	Sin fosforilar. Conformación inactiva. Sin ligando.
2fgi <sup>371</sup>	FGFR1 (Flg)	2.50	456-463 (486-490) (501-504) (580-593) (646, 658) (763-765)	Humano	Unido a PD173074. Conformación inactiva.
1r1w <sup>416</sup>	HGFR	1.80	1050-1360 (1049) (1099-1103) (1231-1244) (1286-1290)	Humano	Autoinhibido. Sin fosforilar. Sin ligando
1r0p <sup>416</sup>	HGFR	1.80	1050-1360 (1049) (1099-1103) (1231-1244) (1286-1290)	Humano	Unido al alcaloide K252 (similar a la staurosporina).
1fvr <sup>417</sup>	TIE-2 (Tek)	2.20	831-1121 (798-812) (861-866) (996-999) (1122-1124)	Humano	Autoinhibido. Sin fosforilar. Sin ligando
1p4o <sup>418</sup>	IGFR-1	1.50	948-1255 (943-947) (1256-1264)	Humano	Inactivo. Sin ligando
1jqh <sup>419</sup>	IGFR-1	2.10	984-1286 (979-983) (1098-1102) (1158-1164)	Humano	Unido a un análogo de ATP y fosforilado.
1m7n <sup>420</sup>	IGFR-1	2.70	974-1286	Humano	Inactivo. Sin ligando
1qcf <sup>421</sup>	Hck	2.00	80-531 (76-79)	Humano	Unido a PP1/AGL1872
1opj <sup>422</sup>	c-Abl	1.75	243-529 (242) (530-534)	Ratón	Unido a STI-571. Inactivo
1iep <sup>423</sup>	c-Abl	2.10	225-498 (223-224) (499-515)	Ratón	Unido a STI-571. Inactivo
1fpu <sup>424</sup>	c-Abl	2.40	225-499 (223-224) (500-515)	Ratón	Unido al inhibidor PRC

1m52 <sup>423</sup>	c-Abl	2.60	232-502 (223-231)(503-515)	Ratón	Unido a PD173955
1mqb <sup>425</sup>	EphA2	2.30	605-887 (568-604) (635-637) (761-777) (888-900)	Humano	Conformación inactiva. Autoinhibido.
1luf <sup>426</sup>	MuSK	2.05	560-860 (526-559) (672-693) (757-760) (861-868)	Rata	Sin ligando.
1opl <sup>422</sup>	c-Abl	3.42	81-531 (1-80) (532-537)	Humano	Unido a pirido[2,3- <i>d</i> ]pirimidina
1opk <sup>422</sup>	c-Abl	1.80	83-531 (40-82) (532-534)	Ratón	Unido a pirido[2,3- <i>d</i> ]pirimidina
1byg <sup>427</sup>	Csk (C-SRC)	2.40	187-448 (173-186) (228-229) (259-260) (336-347) (449-450)	Humano	Unido a staurosporina
1k3a <sup>428</sup>	IGFR-1	2.10	981-1256 (1069-1076)	Humano	Unida a ATP y un péptido sustrato.
1i44 <sup>429</sup>	IGFR-1 mutado	2.40	986-1283 (978-985) (1155-1170)	Humano	Unida a ATP y un péptido sustrato.
1irk <sup>430</sup>	IGFR-1	2.10	981-1283 (978-980) (1005-1007)	Humano	Sin ligando. Conformación inactiva
1k9a <sup>431</sup>	Csk (C-SRC)	2.50	6-450 (1-5) (341-346)	Rata	Sin ligando. Conformación inactiva
1qpe <sup>432</sup>	Lck	2.00	231-501 (502-509)	Humano	Unido a inhibidor PP2. Conformación activa
3lck <sup>433</sup>	Lck	1.60	231-501	Humano	Conformación activa
1qpd <sup>432</sup>	Lck	2.00	231-502 (503-509)	Humano	Unido a staurosporina. Conformación activa
1qpc <sup>432</sup>	Lck	1.60	231-502 (503-509)	Humano	Unido a AMP-PNP. Conformación activa.
1qpj <sup>432</sup>	Lck	2.20	236-501	Humano	Unido a staurosporina
1p14 <sup>434</sup>	IGFR-1 mutado	1.90	978-1283 (1153-1171)	Humano	Sin ligando.
1ir3 <sup>435</sup>	IGFR-1	1.90	981-1283	Humano	Unido a un análogo de ATP. Conformación activa.
1rqq <sup>436</sup>	IGFR-1 mutado	2.60	907-1283 (978-986)	Humano	Conformación activa.
1m14 <sup>401</sup>	EGFR	2.60	672-996 (666-671) (961-978) (997-998)	Humano	Sin ligando. Conformación inactiva.
1m17 <sup>401</sup>	EGFR	2.60	672-995 (666-671) (965-998)	Humano	Unido a OSI-774. Próximo a conformación activa.
1jpa <sup>437</sup>	Neural kinase (EphB2)	1.91	602-900 (595-601) (651-653) (774-796) (901-906)	Ratón	Autoinhibido. Sin fosforilar
1k2p <sup>438</sup>	Quinasa de Bruton (Btk)	2.10	397-654 (655-659)	Humano	Sin ligando. Sin fosforilar
1ksw <sup>439</sup>	c-src (p60-src) mutada	2.80	84-533 (82-83)	Humano	Unida a N6-bencil ADP. Conformación activa.
1fmk <sup>440</sup>	c-src (p60-src)	1.50	86-863 (410-423)	Humano	Conformación inactiva.
2ptk <sup>441</sup>	c-src (p60-src)	2.35	83-528	Pollo	Conformación inactiva
1mp8 <sup>425</sup>	Fak	1.60	414-686 (406-413) (445-446) (565-583)	Humano	Unido a ADP. Conformación no activa.

Res (Å): resolución (todas han sido resueltas por rayos-X).

De todas estas quinasas, c-Kit, Flt3, VEGFR-2 y FGFR-1 presentan un *blast score* con *SWISS-MODEL Blast* inferior a 1E-40. A continuación se encuentran HGFR, TIE-2 e IGFR, con un *blast score* comprendido entre 1E-40 y 1E-30.

Las listadas a continuación presentan un *blast score* inferior a  $1E-30$ . Se analiza en detalle la inclusión o no de estas siete primeras quinasas (c-Kit, Flt3, VEGFR-2, FGFR-1, HGFR, TIE-2 e IGFR). Para cada una de ellas, se selecciona una entrada PDB en función de su resolución, origen y la presencia de inhibidores (con la quinasas tanto en conformación activa como inactiva) y se calcula el porcentaje de identidad (Tabla 5.15).

**Tabla 5.15.** Entradas PDB candidatas seleccionadas con un *blast score* <  $1E-30$ .

Entrada PDB	% Identidad frente a PDGFR- $\beta$	% Identidad de PDGFR- $\beta$ frente a la quinasas	Presencia de ligando
c-Kit: 1t45	45.5	58.6	NO
Flt-3: 1rjb	43.0	60.0	NO
VEGFR: 1vr2	38.8	54.5	NO
FGFR-1: 2fqi	36.9	52.3	SI (PD173074)
HGFR: 1rop	30.4	40.0	SI (alcaloide K252)
TIE-2: 1fvr	31.2	38.9	NO
IGFR-1: 1irk	28.0	36.0	NO

El porcentaje de identidad difiere al expresarlo frente a PDGFR- $\beta$  o la quinasas en cuestión ya que se expresa relativo a la longitud de la secuencia de referencia.

De las siete estructuras mostradas en la Tabla 5.15, se decide tomar como plantillas para la modelización aquellas con una identidad superior al 40%, cuyas regiones características se listan en la Tabla 5.16.

**Tabla 5.16.** Regiones características de las cuatro quinasas seleccionadas como plantillas.

	Entrada SWISS-PROT	Residuos PDB	Región TM	Dominio Quinasas	Sitio de unión del nucleótido	Unión del ATP	Sitio Activo	Dominio KID
c-Kit	P10721	547-935	544-580	589-937 (349)	595-603	Lys623	Asp792	694-753*
Flt-3	P36888	572-947	544-563	610-943 (334)	616-624	Lys644	Asp811	707-782
VEGFR	P35968	820-1168	765-789	834-1162 (329)	840-848	Lys868	Asp1028	940-989
FGFR-1	P11362	460-761	377-397	478-767 (290)	484-492	Lys514	Asp623	No tiene

\*en el PDB 1t45 no aparece, sino que está modelado mediante un vector inserto de una Thr-Ser

El alineamiento múltiple de las cuatro plantillas frente a la secuencia de PDGFR- $\beta$  se realiza con el programa MOE (apartado 1.5.2.1), utilizando una matriz de sustitución BLOSUM-62 (apartado 1.5.2.2) e imponiendo una penalización por apertura y extensión de *gap* de 7 y 1, respectivamente. El resto de parámetros de alineamiento múltiple se dejan por defecto. El alineamiento resultante se refina manualmente, asegurándose que los residuos y dominios conservados se encuentran debidamente alineados.<sup>408,410</sup>

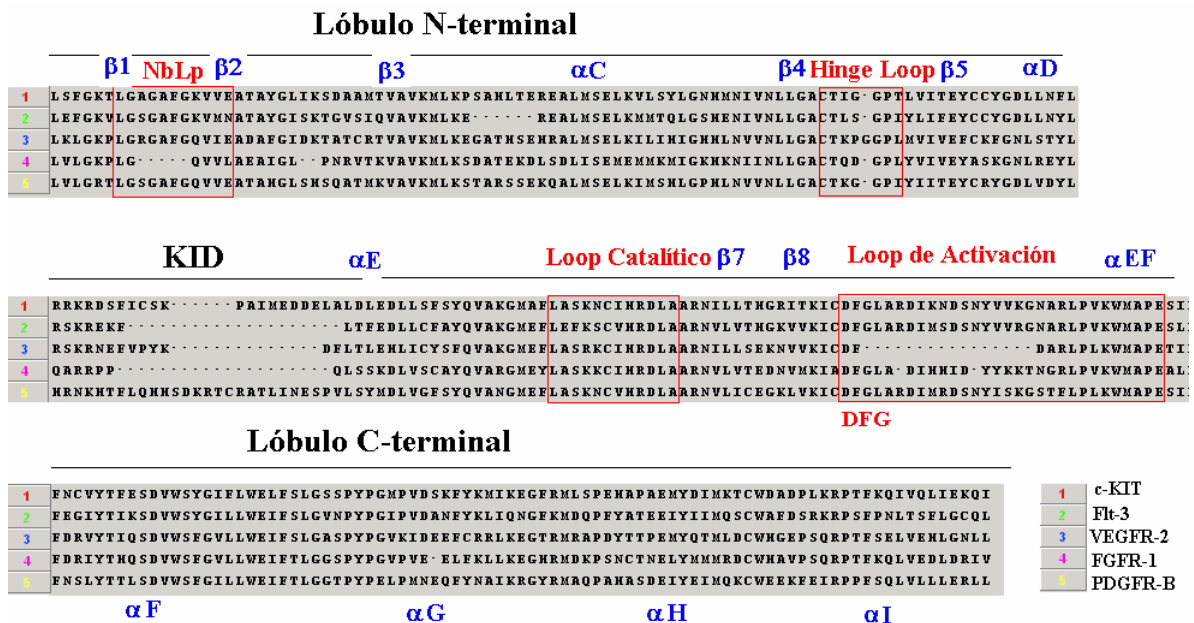
La modelización del dominio KID (residuos 695-798) es conflictiva. Aunque presenta un elevado grado de conservación dentro de esta familia de quinasas<sup>408</sup>, debido a su elevada flexibilidad, su resolución estructural es dificultosa, no encontrándose resuelto en su totalidad en las estructuras cristalizadas de las plantillas. Se encuentran residuos en las plantillas 1t45 y 1vr2 que sirven de esqueleto para la modelización de los segmentos comprendidos entre los residuos Arg<sup>695</sup>-His<sup>704</sup> y Arg<sup>788</sup>-Leu<sup>798</sup>.

Por otro lado, abordar la modelización de un segmento de 104 residuos no es factible mediante métodos *ab initio*. Normalmente, en los modelos por homología construidos para estas TKs con dominios KID se evita su inclusión o bien se introduce un espaciador que conecte los segmentos de los extremos.<sup>410,442</sup>

Se consideran dos opciones para la modelización del dominio KID:

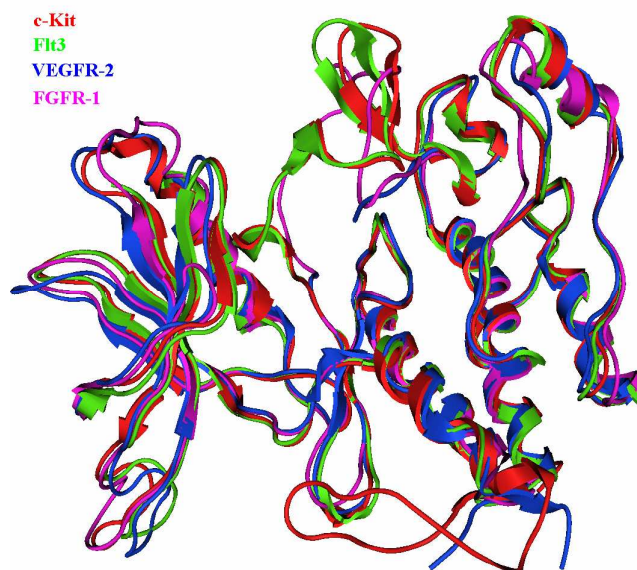
- Inclusión de los segmentos Arg<sup>695</sup>-His<sup>704</sup> y Arg<sup>788</sup>-Leu<sup>798</sup> unidos a partir de un *linker* imaginario compuesto por los tres residuos situados a partir del extremo C-terminal de la His<sup>704</sup> (Ser<sup>705</sup>-Asp<sup>706</sup>-His<sup>707</sup>) y los tres residuos anteriores al extremo N-terminal de Arg<sup>788</sup> (Arg<sup>785</sup>-Thr<sup>786</sup>-Cys<sup>787</sup>). Estos seis residuos se modelarán *ab initio*, actuando en el modelo como mero espaciador. En adelante, se denomina a dicho segmento como dominio KID (aunque en realidad no contiene los 104 residuos).
- Sin incluir ningún residuo de la secuencia comprendida entre la His<sup>704</sup> y la Ser<sup>795</sup>, ya que en el caso anterior el segmento introducido es un artefacto.

Los alineamientos obtenidos en los dos casos se denominan MODEL\_KID y MODEL\_SIN\_KID, respectivamente. En la Figura 5.19 se muestra el alineamiento final MODEL\_KID, destacándose los dominios importantes en la actividad quinasa (apartado 2.4), altamente conservados así como la situación de los elementos de estructura secundaria.



**Figura 5.19.** Alineamiento final considerado de la secuencia del dominio quinasa de PDGFR-β (5) frente a c-Kit (1), Flt-3 (2), VEGFR-2 (3) y FGFR-1 (4).

En la Figura 5.20 se muestra la superposición estructural de las cuatro plantillas.



**Figura 5.20.** Superposición de las cuatro estructuras plantilla consideradas: c-Kit (*rojo*), Flt-3 (*verde*), VEGFR-2 (*azul*) y FGFR-1 (*rosa*). La RMSD de superposición de los átomos de la cadena principal de estas estructuras es: 0.76 (Flt-3), 1.91 (VEGFR-2), 1.80 (FGFR-1) frente a c-Kit; 1.91 (VEGFR-2), 1.87 (FGFR-1) frente a Flt-3; 2.38 (FGFR-1) frente a VEGFR-2.

Cada una de estas cuatro plantillas se acondiciona para la modelización. Para ello se extrae exclusivamente el dominio quinasa de la cadena A de cada fichero *pdb*, eliminándose también las aguas estructurales e inhibidores.

El modelo de PDGFR- $\beta$  se construye empleando la técnica de modelización por satisfacción de restricciones espaciales implementada en MODELLER versión 6.2 (apartado 1.5.3). Se describen diversos aspectos considerados:

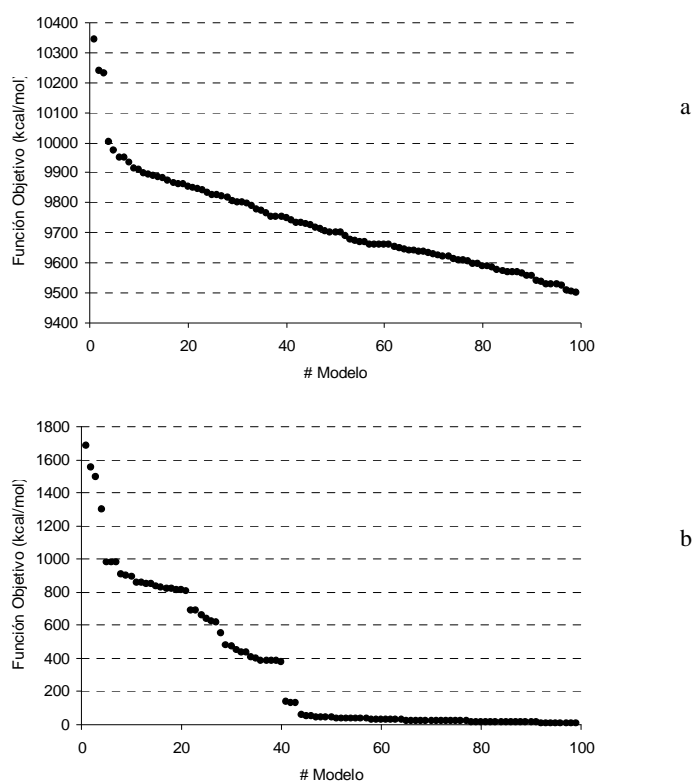
- Para cada uno de los dos posibles alineamientos (MODEL\_KID y MODEL\_SIN\_KID) se generan en principio 100 modelos, seleccionándose aquel de menor valor de función pseudoenergética.
- MODELLER requiere que se indiquen explícitamente los puentes disulfuro. En principio y según la información de la estructura primaria de que se dispone, no existe ninguno en el dominio quinasa de PDGFR- $\beta$ .
- No se consideran restricciones especiales: presencia de prolinas *cis* ( $\omega=0^\circ$ ) o motivos de estructura secundaria. Como se observa en los modelos finales obtenidos, la no imposición de restricciones para las hélices  $\alpha$  o láminas  $\beta$  no impide que éstos se modelen según la topología esperable por comparación con las restantes quinastas.
- Las coordenadas transferidas corresponden al promedio de todas las plantillas. Tras dicha transferencia, se realiza una randomización de coordenadas XYZ de todos los átomos con una desviación de  $\pm 4$  Å.
- En la optimización de cada uno de los modelos se aplica el *Variable Target Function Method* (VTFM) con el *schedule* más exhaustivo de los siete disponibles. A continuación se realiza un refinamiento mediante *Simulated Annealing* con una rampa ascendente de 150 a 1300K y una rampa descendente de 1300K a 300K.



- Las coordenadas de los *loops* de la estructura proceden del promedio de varias plantillas, quedando muy pocos átomos en todo el modelo con coordenadas indefinidas tras la transferencia. Por ello, se decide modelar *ab initio* únicamente el dominio KID insertado, comprendiendo un segmento de 10 residuos. A partir del mejor modelo del alineamiento MODEL\_KID obtenido íntegramente por homología se calculan 100 modelos.

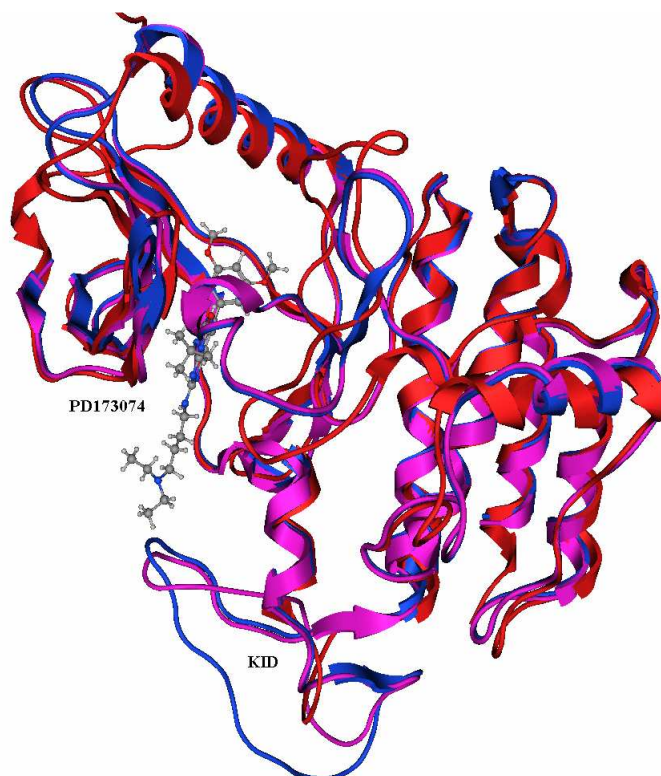
Los modelos finales obtenidos son:

- **MODEL\_KID.** En el gráfico de la Figura 5.21.a se observa el perfil de la función pseudoenergética en orden descendente para los 100 modelos generados a partir del alineamiento MODEL\_KID. El promedio de la RMSD de superposición de todos ellos es de 1.11 Å para todos los átomos de la cadena principal. A partir del mejor de estos modelos, se modela *ab initio* el dominio KID, mostrándose en la Figura 5.21.b el perfil de función pseudoenergética de los 100 modelos requeridos, que presentan una RMSD promedio de 1.31 Å.



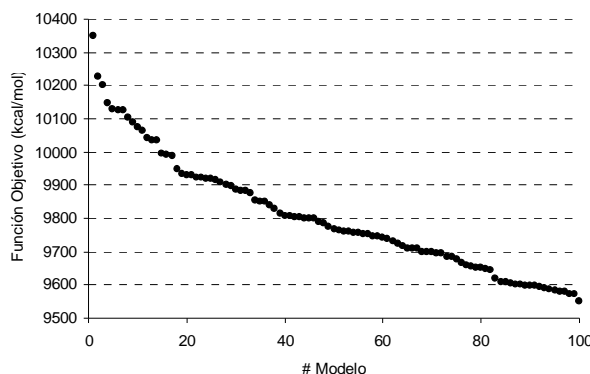
**Figura 5.21.** Perfil de la función objetivo (kcal/mol) pseudoenergética de MODELLER para (a) los 100 modelos de MODEL\_KID y (b) los 100 modelos del dominio KID generados *ab initio* a partir del mejor modelo de la figura anterior.

En la Figura 5.22 se ilustra la superposición de las estructuras con el dominio KID de menor y mayor función pseudoenergética (*rosa* y *azul*, respectivamente) frente a la estructura de FGFR-1 (*rojo*) con el PD173074 co-cristalizado. El dominio KID queda en ambos casos muy alejado de la cavidad de unión, por lo que seguramente su conformación no afectará especialmente al *docking*. Se escoge el modelo de menor función pseudoenergética.



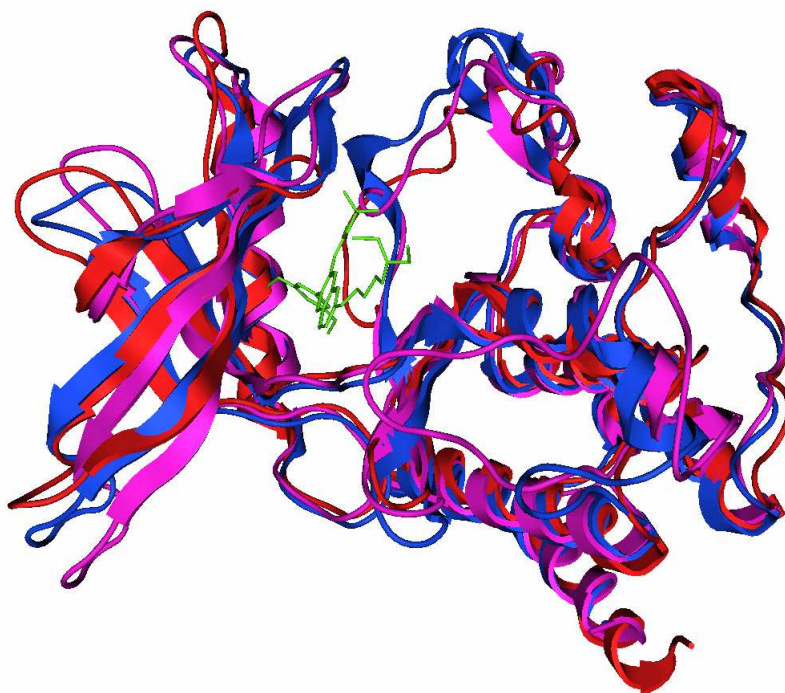
**Figura 5.22.** Superposición de los modelos derivados de MODEL\_KID con el dominio KID modelado *ab initio* de menor (*azul*) y mayor (*rosa*) función pseudoenergética frente a FGFR-1 (*rojo*).

- **MODEL\_SIN\_KID.** Se muestra en la Figura 5.23 el perfil de la función objetivo pseudoenergética para los 100 modelos generados a partir del alineamiento MODEL\_SIN\_KID. Estos modelos presentan una RMSD promedio de 1.20 Å.



**Figura 5.23.** Perfil de la función objetivo (kcal/mol) pseudoenergética de MODELLER para los 100 modelos generados a partir del alineamiento MODEL\_SIN\_KID.

Finalmente, se muestra en la Figura 5.24 la superposición de los modelos finales obtenidos MODEL\_KID (*rosa*) y MODEL\_SIN\_KID (*azul*) frente a la estructura 2fgi de FGFR-1 (*rojo*) con el PD173074 (*verde*). La RMSD de superposición de los átomos de la cadena principal es de 4.02 (MODEL\_KID) y 4.01 (MODEL\_SIN\_KID) frente a 2fgi y de 1.50 entre MODEL\_KID y MODEL\_SIN\_KID.



**Figura 5.24.** Superposición de los modelos MODEL\_KID (*rosa*) y MODEL\_SIN\_KID (*azul*) frente a la estructura 2fgi de FGFR-1 (*rojo*). Se muestra el PD173074 (*verde*).

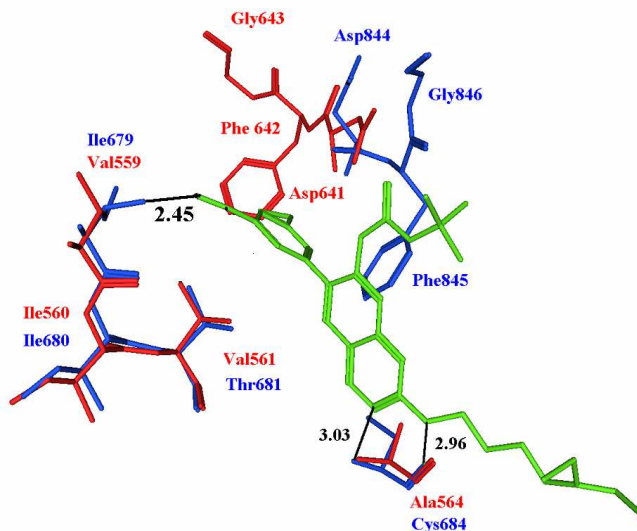
### 5.3.2. Predicción del modo de unión para PD173074 en PDGFR

Además de inhibir a FGFR y VEGFR-2 en el orden nanomolar, PD173074 (Figura 5.1) inhibe también a PDGFR- $\beta$  en el orden micromolar<sup>371</sup>. Según estudios SAR, se ha demostrado que esta selectividad frente a FGFR y VEGFR-2 se obtiene gracias a la derivatización 3,5-dimetoxi en el anillo de fenilo.

En FGFR, el nitrógeno N-3 de PD173074 actúa como aceptor del puente de hidrógeno correspondiente con el nitrógeno amídico de Ala<sup>564</sup> (2.76 Å) y el nitrógeno del grupo butilamino actúa como dador con el oxígeno del carbonilo de Ala<sup>564</sup> (2.54 Å). En PDGFR- $\beta$ , el residuo correspondiente a Ala<sup>564</sup> es Cys<sup>684</sup>. Se ha propuesto<sup>346</sup> que la mayor selectividad de PD173074 frente a FGFR y VEGFR se debe a la presencia de dos residuos valina (Val<sup>559</sup> y Val<sup>561</sup> en FGFR) cuyas cadenas laterales conforman una cavidad hidrofóbica en la que se sitúa este grupo dimetoxifenil. En PDGFR- $\beta$ , la Val<sup>559</sup> corresponde a Ile<sup>679</sup> y la Val<sup>561</sup> a Thr<sup>681</sup>. Dado el mayor tamaño de las cadenas laterales de estos aminoácidos, PD173074 no se acomodaría tan bien en esta cavidad.<sup>346</sup> Además, la conformación de la triada Asp<sup>641</sup>-Phe-Gly al inicio del *loop* de activación podría contribuir a la selectividad de PD173074 por FGFR-1 frente a PDGFR- $\beta$ , pese a ser una secuencia conservada.<sup>371</sup> Otro aminoácido de PDGFR- $\beta$  implicado en la unión de ATP es la Cys<sup>843</sup>, correspondiente a la Ala<sup>640</sup> en FGFR.

En un intento de predecir el modo de unión de PD173074 en PDGFR- $\beta$ , se superponen los dos modelos obtenidos (MODEL\_KID y MODEL\_SIN\_KID) frente a la estructura de FGFR co-cristalizada con PD173074 (2fgi, cadena A). En la Figura 5.25 se detalla la orientación de PD173074 (*verde*) en el sitio activo de MODEL\_KID (*azul*) superpuesto frente a 2fgi (*rojo*). Se muestran los principales residuos implicados. Los dos posibles puentes de hidrógeno de la Cys<sup>684</sup> con PD173074: entre el carbonilo y la cadena lateral (2.96 Å) y entre el nitrógeno amídico y el nitrógeno N-3 de PD173074 se encuentran a distancias de 2.96 y 3.03 Å, respectivamente.

Un metilo del residuo Ile<sup>679</sup> dista 2.45 Å del metilo del grupo metoxi de PD173074, mientras que el metilo más próximo de la correspondiente Val<sup>559</sup> de FGFR-1 se encuentra a 3.60 Å. Además, la conformación de la tríada Asp-Phe-Gly está mucho más próxima en el modelo obtenido para PDGFR (principalmente el fenilo de Phe, que en la perspectiva de la figura se observa como si contactasen, aunque no es así). Una situación muy similar se reproduce para MODEL\_SIN\_KID. En este modelo, la menor distancia de Ile<sup>679</sup> al metilo de PD173074 es ligeramente superior (3.28 Å).

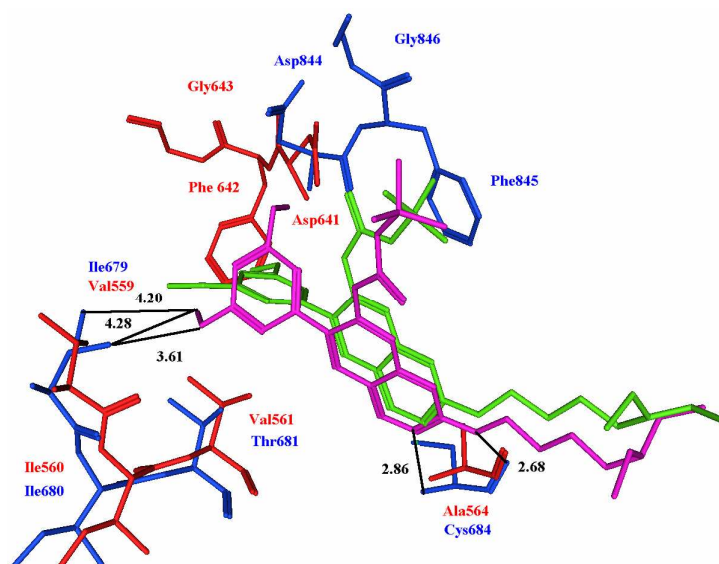


**Figura 5.25.** Unión de PD173074 (verde) a FGFR-1 (rojo) y la superposición del modelo MODEL\_SIN\_KID (azul).

Antes de realizar un *docking* se minimiza el conjunto del modelo de PDGFR correspondiente (MODEL\_KID y MODEL\_SIN\_KID) y PD173074 para eliminar los *bad contacts*, especialmente con Phe<sup>845</sup>. Se añaden todos los hidrógenos y se relaja el sistema con el *force field* MMFF94 usando tres métodos consecutivos (*Steepest Descent*, *Conjugate Gradient* y *Truncated Newton*) para alcanzar una optimización energética con un gradiente de RMS inferior a 0.05.

En la Figura 5.26 se muestra la situación final tras la minimización para el mismo modelo MODEL\_KID (azul) frente a 2fgi (rojo). Se observa el cambio de orientación del metilo en la nueva conformación de PD173074 (rosa) frente a la conformación tras la superposición (verde) alejándose de Ile<sup>679</sup>, la apertura del Phe<sup>845</sup> y el acortamiento de las distancias de puente de hidrógeno con la Cys<sup>684</sup>.

Lo mismo sucede para MODEL\_SIN\_KID, aunque el alejamiento de Ile<sup>679</sup> no es tan notable (3.64 Å *versus* los 3.28 Å anteriores) como en el mostrado en la Figura 5.26.



**Figura 5.26.** Estructura de PD173074 (*rosa*) minimizada en MODEL\_KID (*azul*) superpuesta sobre la estructura de 2fgi (*rojo*) con PD173074 (*verde*).

Una vez minimizados los dos sistemas, se adecuan como *input* para los dos programas (AUTODOCK y GOLD) según el proceso descrito en el apartado 5.1.2. En AUTODOCK, se prueban tamaños de malla de  $61 \times 61 \times 61$  y  $101 \times 101 \times 101$ . El PD173074 se parametriza con las cargas Gasteiger-Marsili, calculándose los potenciales de afinidad atómica estándar y el conjunto adicional (halógenos, nitrógeno aceptor de puente de hidrógeno...).

En la Tabla 5.17 se muestran las distancias para las dos interacciones de puente de hidrógeno de PD173074 con Cys<sup>684</sup> resultantes para la conformación de menor *docked energy* en los sucesivos *dockings* con los cuatro protocolos:

- i) AUTODOCK con tamaños de malla de  $61 \times 61 \times 61$
- ii) AUTODOCK con tamaños de malla de  $101 \times 101 \times 101$
- iii) GOLD con GOLDScore y
- iv) GOLD con CHEMScore

A modo orientativo, se muestra en la Tabla 5.17 la RMSD de superposición de la conformación resultante frente a la conformación inicial de PD173074, correspondiente a la inicialmente minimizada en la cavidad de PDGFR- $\beta$ . En la Tabla 5.17 se observa cómo los resultados de AUTODOCK con una malla pequeña predicen una interacción conservada para los dos modelos (distancias de 2.6, 3.0 y 4.0, 4.3, respectivamente), aunque al aumentar el tamaño de la malla a  $101 \times 101 \times 101$ , las conformaciones se apartan de la cavidad de unión. Por otra parte, GOLDScore orienta las conformaciones prácticamente cómo han quedado tras la minimización para el modelo MODEL\_KID, aunque para MODEL\_SIN\_KID la conformación obtenida está totalmente *misdocked*. Para CHEMScore, las conformaciones están totalmente *misdocked*.

**Tabla 5.17.** Distancias de interacción (Å) para los dos puentes de hidrógeno entre la Cys<sup>684</sup> y PD173074 (**17**, Figura 5.1) y sus desviaciones estándar (paréntesis) obtenidas para las posiciones inicialmente situadas en el sitio de unión en el *docking* frente a los dos modelos obtenidos para PDGFR-β y los cuatro protocolos. RMSD (Å) y sus desviaciones estándar (paréntesis) respecto a la conformación inicial minimizada en la cavidad de unión.

		MODEL_KID	MODEL_SIN_KID
AUTODOCK	D (C=O a NH)	2.6 (0.0)	4.0 (0.4)
Mapas de afinidad adicionales	D (NH a N3)	3.0 (0.2)	4.3 (0.3)
61×61×61	RMSD	1.9 (0.0)	3.1 (0.4)
AUTODOCK	D (C=O a NH)	10.5 (6.6)	12.3 (6.5)
Mapas de afinidad adicionales	D (NH a N3)	10.7 (5.8)	12.3 (5.6)
101×101×101	RMSD	5.2 (3.2)	9.2 (4.3)
GOLD	D (C=O a NH)	2.8 (0.2)	5.4 (4.0)
GOLDScore	D (NH a N3)	2.9 (0.2)	3.8 (1.6)
	RMSD	0.8 (0.1)	4.8 (3.7)
GOLD	D (C=O a NH)	6.6 (0.8)	7.1 (0.1)
CHEMScore	D (NH a N3)	4.8 (1.1)	8.4 (1.0)
	RMSD	9.0 (3.1)	5.7 (0.1)

*D(C=O a NH)*: distancia entre el oxígeno del carbonilo de la Cys<sup>684</sup> y el grupo amino de la cadena lateral de PD173074. *D (NH a N3)*: distancia entre el nitrógeno amídico de la Cys<sup>684</sup> y el nitrógeno N3 de PD173074 (**17**, Figura 5.1).

### 5.3.3. Resultados del *docking* ciego de PD173074 en PDGFR-β

Finalmente se realiza un *docking* ciego de PD173074 en los dos modelos de PDGFR-β con AUTODOCK y una malla de 181×181×181. En la Tabla 5.18 se muestran las distancias entre los átomos implicados en los posibles puentes de hidrógeno con Cys<sup>684</sup>, así como la RMSD de superposición frente a la estructura inicialmente minimizada en la cavidad de unión. Ninguno de los conformémeros encuentra o se orienta adecuadamente en la cavidad de unión. Esto es predecible en base a los resultados mostrados en la Tabla 5.17, en los que al aumentar el tamaño de la malla, los ligandos quedan mal orientados.

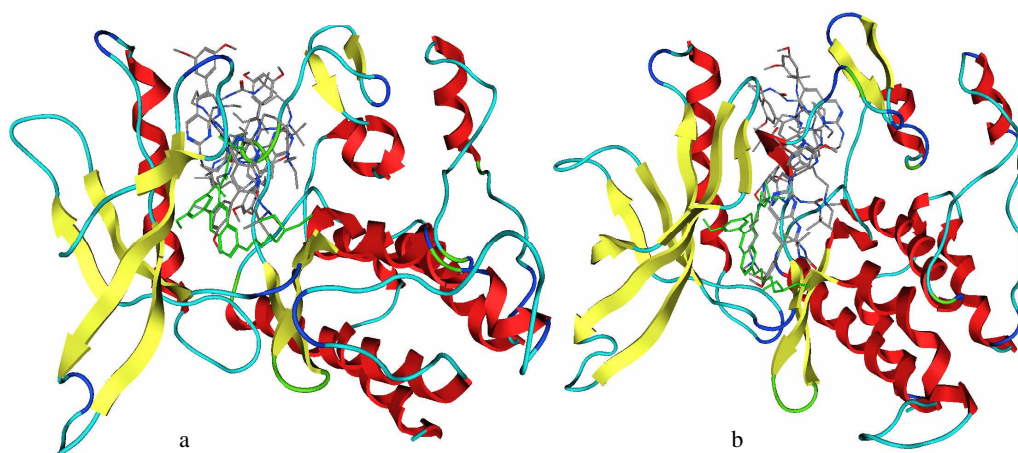
**Tabla 5.18.** Distancias de interacción (Å) para los dos puentes de hidrógeno entre la Cys<sup>684</sup> y PD173074 (**17**, Figura 5.1) y sus desviaciones estándar (paréntesis) obtenidas para las posiciones inicialmente situadas en la superficie de PDGFR-β en el *docking* con AUTODOCK frente a los dos modelos obtenidos. RMSD (Å) y sus desviaciones estándar (paréntesis) respecto a la conformación inicial minimizada en la cavidad de unión.

		MODEL_KID	MODEL_SIN_KID
AUTODOCK	D (C=O a NH)	8.3 (5.7)	16.5 (4.3)
Mapas de afinidad adicionales	D (NH a N3)	9.5 (5.4)	18.3 (4.1)
61×61×61	RMSD	5.5 (3.5)	11.8 (2.8)

*D(C=O a NH)*: distancia entre el oxígeno del carbonilo de la Cys<sup>684</sup> y el grupo amino de la cadena lateral de PD173074. *D (NH a N3)*: distancia entre el nitrógeno amídico de la Cys<sup>684</sup> y el nitrógeno N3 de PD173074 (**17**, Figura 5.1).

En la Figura 5.27 se muestran las cinco conformaciones de mínima energía obtenidas en el *docking* frente a MODEL\_KID (a) y MODEL\_SIN\_KID (b). De manera análoga a lo sucedido frente a FGFR (apartado 5.1.3), se observa cómo las conformaciones tienden a posicionarse en la cavidad del *loop* de activación.





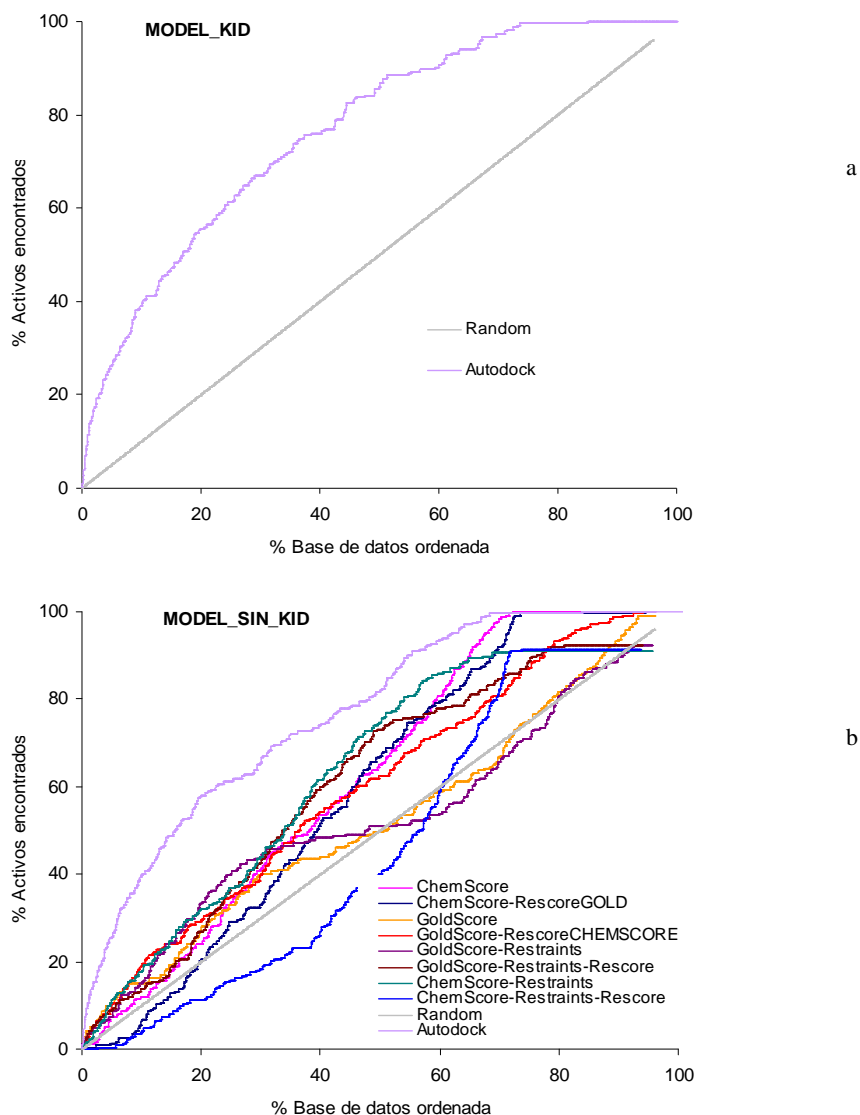
**Figura 5.27.** Cinco primeros conformeros obtenidos en el *docking* ciego de PD173074 frente a MODEL\_KID (a) y MODEL\_SIN\_KID (b). La conformación minimizada en la cavidad de unión se muestra en verde.

#### 5.3.4. Cribado virtual en PDGFR- $\beta$

En el cribado virtual frente a PDGFR- $\beta$  se utiliza Base\_Maybridge (apartado 5.2.4). Del *pool* de activos ACTIV\_1, compuesto por 288 inhibidores de FGFR, se extraen aquellos compuestos cuya actividad ha sido determinada frente a PDGFR- $\beta$  y que presentan una  $IC_{50}$  inferior a  $50\mu\text{M}$ . En total, se contabilizan 234 compuestos inhibidores de PDGFR- $\beta$ : 70 7-alquilurea pirido[2,3-*d*]pirimidinas, 75 pirido[2,3-*d*]pirimidinas, 24 naftiridin-2(1*H*)-onas, 38 indolin-2-onas y 27 1-fenilbenzimidazoles. Sus estructuras y valores de  $IC_{50}$  frente a PDGFR se muestran en el Anexo 2.

Así, la base de datos utilizada consiste de 234 activos y 4342 inactivos (4576 compuestos), presentando un factor de enriquecimiento máximo de 19.6 (4576/234). En adelante, se denomina Base\_Maybridge\_234.

Se realiza el cribado virtual con el programa AUTODOCK, frente a MODEL\_KID y MODEL\_SIN\_KID, y seguidamente, con el programa GOLD únicamente frente a MODEL\_SIN\_KID. En la Figura 5.28 se muestran las curvas de enriquecimiento obtenidas en el *docking* frente a MODEL\_KID (a) y MODEL\_SIN\_KID (b).



**Figura 5.28.** Curvas de enriquecimiento obtenidas con los protocolos de *docking* frente PDGFR para las estructuras (a) MODEL\_KID y (b) MODEL\_SIN\_KID en el cribado de Base\_Maybridge\_234. La línea gris representa los valores obtenidos si los activos se seleccionasen aleatoriamente.

Para MODEL\_KID, los valores de *ef* obtenidos con AUTODOCK en el primer 2%, 5% y 10% de base cribada son 8.5, 5.5 y 3.9, respectivamente. En la Tabla 5.19 se muestran los factores de enriquecimiento (*ef*) en el primer 2%, 5% y 10% de base de datos Base\_Maybridge\_234 cribada frente al MODEL\_SIN\_KID.

Como en el caso de EGFR, AUTODOCK es el protocolo que mejor se comporta en el cribado frente a PDGFR- $\beta$  en términos de enriquecimiento. En este caso, la construcción de un modelo por homología es útil en la discriminación entre activos e inactivos. En casos como CHEMSCORE, la recuperación de activos es inferior a una selección aleatoria. La inclusión de restricciones proporciona un incremento en el factor de enriquecimiento en el caso de CHEMSCORE, aunque no tanto en GOLDScore.

Respecto a los modelos, MODEL\_KID alcanza ligeramente un enriquecimiento superior en el primer 2% de base de datos cribada (8.8 *versus* 7.7), aunque apenas se encuentran diferencias.



**Tabla 5.19.** Factores de enriquecimiento (*ef*) obtenidos en el primer 2%, 5% y 10% de base de datos Base\_Maybridge\_234 cribada frente PDGFR.

	2%	5%	10%
	MODEL_SIN_KID	MODEL_SIN_KID	MODEL_SIN_KID
CHEMSCORE	0.6	1.5	4.0
CHEMSCORE-RESCORE	0.2	0.3	0.5
CHEMSCORE-RESTRAINT	1.9	2.2	1.8
CHEMSCORE-RESTRAINT-RESCORE	0.0	0.1	0.4
GOLDScore	2.8	2.1	1.5
GOLDScore-RESCORE	2.6	2.0	1.9
GOLDScore-RESTRAINT	2.3	1.5	1.5
GOLDScore-RESTRAINT-RESCORE	2.8	1.8	1.4
AUTODOCK	7.7	5.1	4.0

En la Tabla 5.20 se muestran los porcentajes de base de datos muestreada a los que se encuentra el primer representante de cada uno de los cinco quimiotipos contenidos en el *pool* de activos, así como si la orientación de estos compuestos reproduce (A) o no (M) el modo de unión previsible, en función del modo de unión de PD173074 superpuesto. Como en casos anteriores, se presenta la existencia (1) o no (0) de un puente de hidrógeno entre el correspondiente compuesto y la Cys<sup>684</sup> de PDGFR- $\beta$ .

Los 1-fenilbenzimidazoles se recuperan a porcentajes tardíos, con una conformación *misdocked* en todos los casos. Por otra parte, las naftiridinas adoptan con todos los algoritmos una conformación previsiblemente bien orientada hacia la Cys<sup>684</sup>. Los otros dos *scaffolds* con anillos de seis miembros, pirido[2,3-*d*]pirimidinas y 7-alquileura pirido[2,3-*d*]pirimidinas, presentan una mayor variabilidad, adoptando tanto orientaciones activas como *misdocked*. A la vista de los resultados de enriquecimiento y diversidad en la identificación de quimiotipos, AUTODOCK es claramente el protocolo óptimo en este caso.

**Tabla 5.20.** Porcentaje de base de datos cribada frente a PDGFR- $\beta$  para la que al menos un representante activo de cada uno de los cinco *scaffolds* de Base\_Maybridge\_234 es recuperado.

	MODEL_SIN_KID					MODEL_KID				
	PD	Naph	Urea	Indol	Phen	PD	Naph	Urea	Indol	Phen
CHEMSCORE	17.3 M(0)	4.5 A(1)	<b>38.0</b> M(0)	6.2 A(1)	5.9 M(1)					
CHEMSCORE-RESCORE	0.7 A(1)	6.0 A(1)	8.5 M(0)	<b>13.6</b> A(1)	9.6 M(1)					
CHEMSCORE-RESTRAINT	10.2 A(1)	1.9 A(1)	<b>15.7</b> M(1)	0.2 A(1)	5.1 M(0)					
CHEMSCORE-RESTRAINT-RESCORE	2.6 A(1)	11.5 A(1)	6.8 M(1)	<b>14.5</b> A(1)	12.2 M(0)					
GOLDSCORE	<b>13.9</b> A(1)	11.3 A(1)	0.1 A(1)	11.1 M(0)	45.9 M(0)					
GOLDSCORE-RESCORE	6.3 M(0)	2.0 A(1)	0.1 A(1)	0.5 A(1)	<b>27.3</b> M(0)					
GOLDSCORE-RESTRAINT	11.5 M(1)	9.7 A(1)	0.1 A(1)	11.5 A(1)	<b>69.6</b> M(0)					
GOLDSCORE-RESTRAINT-RESCORE	2.7 M(0)	<b>8.6</b> A(1)	5.6 A(1)	0.1 A(1)	1.1 M(0)					
AUTODOCK	0.09 A(1)	0.1 A(1)	0.02 A(1)	1.0 M(1)	<b>28.8</b> M(0)	0.02 M(0)	0.2 M(0)	0.1 A(1)	0.9 A(1)	<b>32.1</b> M(0)
PROMEDIO	7.3	6.2	8.3	6.5	22.8	0.02	0.2	0.1	0.9	32.1

PD: pirido[2,3-*d*]pirimidinas, Naph: naftiridin-2(1*H*)-onas, Urea: 7-alquilurea pirido[2,3-*d*]pirimidinas, Indol: indolin-2-onas y Fenil: 1-fenilbenzimidazoles. Los números en negrita corresponden al peor caso, es decir, el porcentaje de base de datos ordenada necesario para recuperar al menos un representante de cada uno de los *cores*. A y M corresponden a la clasificación de las conformaciones en *active site* y *misdocked*, respectivamente. El número en paréntesis indica si se detecta (1) o no (0) puente de hidrógeno con el residuo Cys<sup>684</sup> de PDGFR- $\beta$  implicado en el modo de unión.

#### 5.4. Comparación del cribado virtual *ligand-based* y *structure-based*

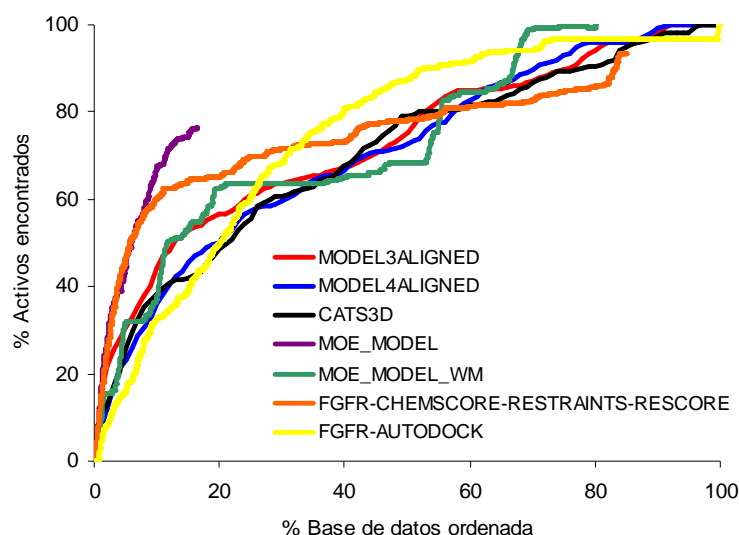
De las tres quinasas estudiadas, es en el cribado retrospectivo frente a FGFR (concretamente con la estructura 2fgi) en el que se alcanzan enriquecimientos mayores. Hay que considerar que el *pool* de activos considerado se recopila focalizado principalmente en compuestos inhibidores de FGFR y que aunque también son activos frente a EGFR y PDGFR- $\beta$ , no se ha realizado el esfuerzo de recopilar un mayor número de activos con un mayor perfil de selectividad hacia estas dos últimas quinasas, lo que podría afectar a los resultados de cribado virtual presentados.

En la Figura 5.29 se comparan los enriquecimientos obtenidos para Base\_ACTIV\_1 con los cinco métodos *ligand-based* seleccionados en el capítulo 4 (apartado 4.10):

- MOE\_MODEL obtenido por refinamiento manual en MOE. MOE\_MODEL\_WM (sin múltiples PPPs).
- Búsqueda con descriptores CATS3D con PD173074 como *focus* y con distancia Manhattan.
- Modelo SQUID derivado de MODEL3ALIGNED.
- Modelo SQUID derivado de MODEL4ALIGNED.

con el cribado virtual *structure-based* frente a FGFR (2fgi) con

- CHEMSCORE-RESTRAINTS-RESCORE.
- AUTODOCK.



**Figura 5.29.** Comparación de las curvas de enriquecimiento obtenidas con las búsquedas de similitud y *docking* frente a FGFR para Base\_ACTIV\_1.

El cribado virtual con los métodos de *docking* se sitúa en el primer porcentaje de base de datos cribada al mismo nivel que el mejor modelo farmacofórico obtenido. En la Tabla 5.21 se indican los porcentajes de base de datos necesarios para identificar al menos un compuesto representativo de cada *scaffold*. Los métodos *ligand-based* identifican el grupo de las indolinonas en posiciones más retrasadas, mientras que en el *docking*, el grupo más retrasado son las anilinoquinazolinonas y fenilbenzimidazoles. En este sentido, existe una complementación deseable entre ambas estrategias.

**Tabla 5.21.** Porcentaje de Base\_ACTIV\_1 cribada para que al menos un representante activo de cada uno de los seis *scaffolds* sea recuperado. Comparación entre métodos *ligand-based* y *structure-based*.

	2fgi					
	PD	Naph	Urea	Indol	Phen	Anil
MODEL3ALIGNED	0.0	0.8	0.5	<b>18.5</b>	3.6	0.1
MODEL4ALIGNED	0.3	0.6	1.1	<b>15.3</b>	5.1	0.1
CATS3D	2.2	5.7	0.0	8.6	<b>19.1</b>	3.4
MOE_MODEL	0.0	0.1	0.0	<b>7.4</b>	3.1	2.4
FGFR CHEMSCORE-RESTRAINTS-RESCORE	0.6	1.0	0.1	0.1	14.8	<b>65.3</b>
FGFR AUTODOCK	0.6	0.7	0.7	0.9	<b>16.4</b>	1.2

PD: pirido[2,3-*d*]pirimidinas, Naph: naftiridin-2(1*H*)-onas, Urea: 7-alkilurea pirido[2,3-*d*]pirimidinas, Indol: indolin-2-onas, Fenil: 1-fenilbenzimidazoles y Anilino: 4-anilinoquinazolinas. Los números en negrita corresponden al peor caso, es decir, el porcentaje de base de datos ordenada necesario para recuperar al menos un representante de cada uno de los *cores*.

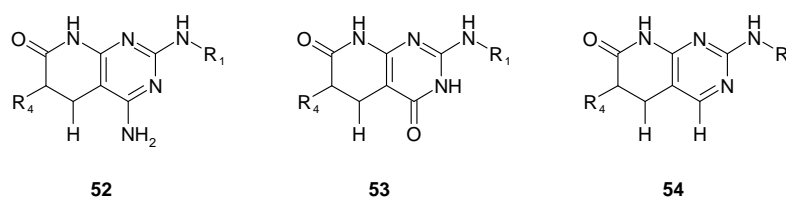
### 5.5. Filtrado prospectivo de las quimiotecas BIB\_Oxo, BIB\_Amino y BIB\_Hidro

En base a las predicciones *ligand-based* (apartado 4.15), se opta por realizar el *docking* de las quimiotecas BIB\_Amino y BIB\_Hidro (71280 compuestos), desestimándose BIB\_Oxo, ya que son estas dos las que presentan mayor *scoring* de similitud.

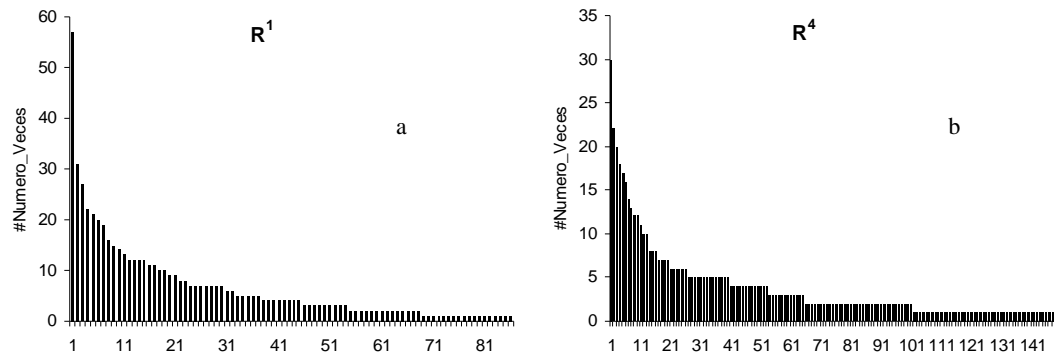
Como estructura de quinasa, se escoge FGFR (2fgi) con el programa GOLD y el protocolo CHEMSCORE-RESTRAINTS-RESCORE, que es el que mejor se comporta desde el punto de vista de enriquecimiento, en lugar de AUTODOCK, que es el que mejor se comporta desde el punto de vista de diversidad de *scaffold* (ya que en este caso el *scaffold* es compartido entre las quimiotecas). Tras el *docking* de estos 71280 compuestos, se seleccionan los 100 primeros de mayor función de *scoring*. De ellos, 78 compuestos comparten el *scaffold* 4-hidro (BIB\_Hidro) y 22 pertenecen a BIB\_Amino.

Se analiza el solapamiento existente entre las selecciones de 100 compuestos identificados separadamente por cada uno de los 6 métodos: MOE\_MODEL, CATS3D-PD173074, CATS3D-SU5402, MODEL3ALIGNED, MODEL4ALIGNED y el *docking* frente a 2fgi. Del total de 600 compuestos, únicamente se solapan 2 compuestos entre MOE\_MODEL y CATS3D-SU5402 así como 28 compuestos ente MODEL3ALIGNED y MODEL4ALIGNED. Así, en total quedan 570 compuestos diferentes.

Además, se analiza la frecuencia de aparición de cada uno de los restos R<sup>1</sup> y R<sup>4</sup> (Figura 5.30) en el total de los 570 compuestos seleccionados (Figura 5.31). Del total de 165 y 216 restos respectivos, los distintos filtrados incluyen 86 y 148 de ellos, respectivamente. Se observa cómo existen restos con una mayor frecuencia de aparición que los restantes.



**Figura 5.30.** Estructura de los sistemas 4-aminopirido[2,3-*d*]pirimidínicos (52), 4-oxopirido[2,3-*d*]pirimidínicos (53) y 4-hidropirido[2,3-*d*]pirimidínicos (54).



**Figura 5.31.** Frecuencia de aparición de los restos (a) R<sup>1</sup> y (b) R<sup>4</sup> de las quimiotecas BIB\_Amino, BIB\_Hidro y BIB\_Oxo en las selecciones realizadas con los métodos *ligand-based* y *structure-based*.

En base a todos estos filtrados y considerando las dificultades sintéticas, se están analizando actualmente aquellos de los 570 candidatos propuestos que deben ser sintetizados en un futuro como potenciales inhibidores de tirosina quinasas.



## Capítulo 6.

### Implementación de un *fingerprint* de interacción

Uno de los problemas que se encuentra en el *docking* es la evaluación de las conformaciones obtenidas de los ligandos. En este sentido, resulta de interés disponer de un método automático que pueda evaluar la calidad de la reproducción de los modos de unión, evitando la inspección visual de los diferentes compuestos.

Recientemente se han diseñado unos *fingerprints* de interacción (*Interaction fingerprints*, IFs) que facilitan la representación y el análisis de las interacciones tridimensionales entre proteína y ligando.<sup>145</sup> Particularmente, se han mostrado muy útiles en el procesado de los resultados de un *docking* y en el cribado virtual. De hecho, una de sus primeras aplicaciones se encuentra en el campo de los inhibidores de proteína quinasas, concretamente la quinasa p38.<sup>443</sup>

Estos métodos codifican los contactos proteína-ligando en cadenas de bits cuya longitud se deriva del número de átomos de la proteína diana presentes en su cavidad de unión. Típicamente, para cada interacción particular (puente de hidrógeno, contacto hidrofóbico vdW...) detectada entre un átomo del ligando y la proteína, se activa (1) el correspondiente bit del átomo o residuo implicado del receptor (Figura 6.1).

Puente de Hidrógeno							Contacto vdW												
...	Leu100				Lys101			...	...	Leu100				Lys101			...		
...	N	...	CD1	CD2	N	...	O	NZ	...	...	N	...	CD1	CD2	N	...	O	NZ	...
...	0	...	0	0	0	...	1	0	...	...	0	...	1	0	0	...	0	0	...

**Figura 6.1.** Representación esquemática de la derivación de un *fingerprint* de interacción.

Las diferentes implementaciones varían en función de cómo definen la cadena de bits, el tipo de interacciones que consideran y cómo las evalúan:

- La propuesta inicial de Deng y colaboradores<sup>145,443</sup> opera a nivel de residuos. Una vez identificados aquellos residuos que conforman el sitio activo, se detectan los contactos hidrofóbicos y los puentes de hidrógeno existentes en el complejo proteína-ligando. Cada residuo del receptor incluye 7 bits asignables a siete posibles tipos de interacciones: i) si está en contacto con el ligando, ii) si participa algún átomo de la cadena principal, iii) si participa algún átomo de la cadena lateral del aminoácido, iv) si se trata de una interacción polar, v) si se trata de una interacción no polar, vi) si el residuo de la proteína actúa como aceptor de puente de hidrógeno y vii) si actúa como dador de puente de hidrógeno. Con ello, la longitud del *fingerprint* equivale al número de residuos definidos como sitio activo multiplicado por siete.
- Muy poco tiempo después, Kelly y Mancera<sup>444</sup> trasladaron el concepto inicial basado en residuos a uno basado en átomos. Tras definir los residuos contenidos dentro de 6 Å de un ligando co-cristalizado, se identifican los contactos por puente de hidrógeno que activarán el bit correspondiente del átomo de la proteína que lo establece. Además, no se limita a definir existencia de contacto (1) o ausencia (0), sino que se pondera la importancia de las interacciones detectadas considerando diversas aproximaciones.

- Mpamhanga y Willet<sup>445</sup> desarrollan una nueva aproximación basada en la idea de que no todos los posibles *leads* tienen porqué presentar el mismo modo de unión que el compuesto de referencia, sino que pueden compartir muchos motivos de interacción similares. Desarrollan tres IFs, cuya longitud viene dada por el número de *heavy atoms* del sitio de unión comprendidos dentro de un radio desde el ligando (típicamente 10 Å).
  - ◆ HIF: basado en puentes de hidrógeno. Una posible interacción por puente de hidrógeno se detecta, activándose el correspondiente bit del átomo del receptor, si el par aceptor-hidrógeno polar se encuentra a una distancia inferior a 3 Å y con un ángulo inferior a 90°.
  - ◆ CIF: basado en contactos hidrofóbicos. Se activa el bit correspondiente del átomo del receptor si está implicado en una interacción por contacto hidrofóbico con el ligando. Ésta se define para cualquier par de átomos diferentes de hidrógeno, que no participen en una interacción por puente de hidrógeno, situados a una distancia inferior a la suma de sus radios de vdW.
  - ◆ CHIF: combinación de los dos anteriores.

Para posibilitar la descripción de múltiples modos de unión, se combina la información de varios complejos cristalográficos en un único *fingerprint* de referencia a partir de un análisis ponderado de la frecuencia de los contactos detectados en varias estructuras de referencia.

Una vez codificada la información, se realiza una búsqueda de similitud frente a un *fingerprint* de referencia, típicamente calculado para un complejo cristalográfico. Así, se comparan los *fingerprints* derivados para cada una de las conformaciones de todos los compuestos de una quimioteca “*dockeada*” (acoplada) frente a la proteína diana con el *fingerprint* de referencia usando los coeficientes de similitud descritos en el apartado 1.10.1. De hecho, este valor de similitud o *scoring* se puede utilizar en el cribado virtual de diversas formas:

- De todas las conformaciones de un ligando concreto, se selecciona aquella que presenta mayor similitud a la conformación de referencia (criterio estructural o de similitud). Posteriormente se ordenan los compuestos (únicamente la conformación óptima escogida) en función de su función de *scoring* del *docking* (criterio energético).
- Se selecciona la conformación óptima de un ligando y los compuestos se ordenan en función del *scoring* de similitud obtenido.
- Se combina el *scoring* de similitud con el *scoring* de *docking* en una única función utilizada tanto para evaluar la mejor conformación de cada ligando como para realizar el ranking de los compuestos.<sup>445</sup>

En este capítulo se presenta una nueva propuesta para codificar los modos de unión de un ligando y proteína en un *fingerprint* basado en el concepto de pares atómicos (apartado 1.6.5). En la siguiente sección se describe su implementación. El método se valida en el cribado virtual frente a tres dianas terapéuticas: i) el receptor  $\alpha$  de estrógeno, ii) FGFR y iii) la transcriptasa reversa del HIV. En los tres casos, los resultados se comparan con los obtenidos aplicando el *fingerprint* CHIF introducido por Mpamhanga, que fue validado originalmente en el cribado virtual retrospectivo de antagonistas del receptor  $\alpha$  de estrógeno.<sup>445</sup>



### 6.1. Descripción del *fingerprint* propuesto: IFbAP

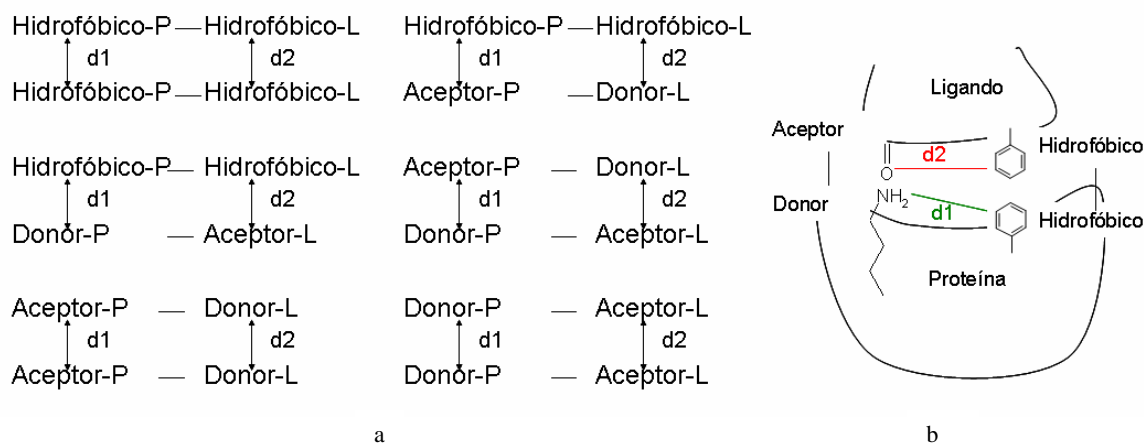
Esta aproximación considera en el *fingerprint* la posición relativa de los pares de parejas de átomos que interactúan, en lugar de la localización absoluta de una interacción.

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. En total existen seis posibles combinaciones de pares tipos de contactos de interacción (Figura 6.2.a).

Para cada par de interacciones detectadas en el complejo, se mide la distancia entre los dos átomos del receptor ( $d_1$ ) y la distancia entre los dos átomos del ligando ( $d_2$ ) (Figura 6.2.b). Cada una de estas distancias incrementa una cuenta en un *bin* dividido en siete rangos, extraídos de la referencia de Mason<sup>279</sup> y que corresponden los rangos de distancias (Å): [0-2.5], [2.5-4], [4-6], [6-9], [9-13], [13-18] y [ $>18$ ]. La posición del bit ocupado (enumerado desde cero) queda definida por el *bin* de ambas distancias (enumerado desde cero), según la ecuación [6.1]:

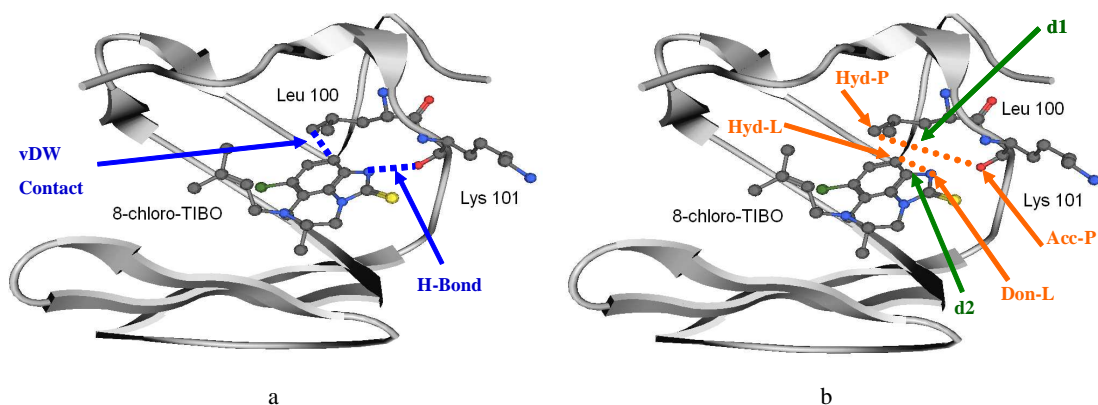
$$\text{Posición del bit} = \text{bin}(d_2) + 7 \cdot \text{bin}(d_1) \quad [6.1]$$

La longitud final del *fingerprint* corresponde al número de combinaciones posibles de pares de tipos de contactos de interacción (seis) y las dimensiones del esquema de *binning* para las distancias del ligando (siete) y las distancias del receptor (siete). Así, el *fingerprint* total se compone de  $6 \times 7 \times 7 = 294$  bits.



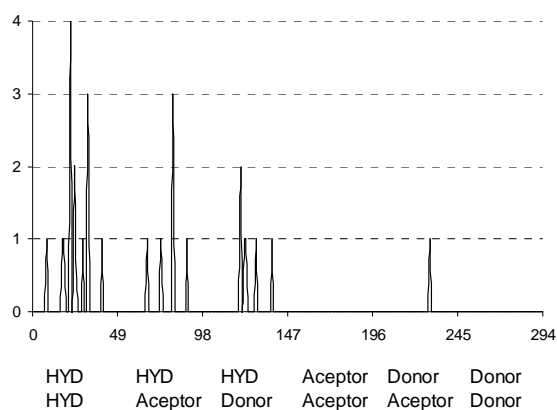
**Figura 6.2.** (a) Seis posibles combinaciones de pares de interacciones que definen un conjunto de 49 bits (7 por cada distancia). (b) Codificación del par de interacciones a partir de las distancias medidas entre los dos átomos de la proteína ( $d_1$ ) y del ligando ( $d_2$ ) que participan.

En la Figura 6.3 se esquematiza el proceso de generación de un *fingerprint* para el caso del complejo de 8-cloro-TIBO con la transcriptasa reversa del HIV-1. Para un par de interacciones detectadas, contacto de vDW entre el átomo CD1 de Leu<sup>100</sup> y un carbono de 8-cloro-TIBO y un puente de hidrógeno entre el oxígeno carbonílico de Lys<sup>101</sup> y el nitrógeno de 8-cloro-TIBO (Figura 6.3.a), se mide las distancias correspondientes entre los dos átomos de la proteína implicados ( $d_1$ ) y entre los dos átomos del ligando ( $d_2$ ) (Figura 6.3.b).



**Figura 6.3.** Esquema de la generación de *fingerprints* para el caso del complejo entre 8-cloro-TIBO y la HIV-1 transcriptasa reversa. (a) Identificación de los contactos existentes. (b) Medida de las distancias implicadas para cada par de interacciones.

En la Figura 6.4 se muestra el correlograma del *fingerprint* obtenido para el complejo 2fgi de FGFR co-cristalizada con PD173074. Como sería esperable, el mayor número de interacciones hidrofóbicas existentes hace que el sector de estos 49 primeros bits obtenga un mayor número de cuentas que, por ejemplo, la única cuenta entre la interacción entre un aceptor-dador, correspondiente a las dos interacciones por puente de hidrógeno entre PD173074 con Ala<sup>542</sup>. En este sentido, se han construido tanto *fingerprints* sin normalizar, como el mostrado en la Figura 6.4, como *fingerprints* normalizados entre 0 y 1, dividiendo por el valor máximo del *fingerprint*.



**Figura 6.4.** Correlograma del *fingerprint* del complejo PD173074 en FGFR (estructura 2fgi)

Este método se implementa en el programa MOE, en lenguaje SVL. Para definir las interacciones, se utiliza la función *pro\_Contacts* de MOE, que utiliza la función de Stickle<sup>446</sup> para definir los puentes de hidrógeno y los contactos hidrofóbicos se determinan con un valor límite de 4.5 Å.

En adelante se denomina a este *fingerprint* IFbAP (*Interaction Fingerprint based on Atom Pairs*).

Como se ha comentado, también se implementa en MOE el *fingerprint* descrito por Mpamhanga<sup>445</sup>, atendiendo fidedignamente a su descripción. De los tres *fingerprints* propuestos por estos autores (HIF, CIF y CHIF), es CHIF el que obtiene mejores enriquecimientos, por lo que únicamente se considera éste. La determinación de los puentes de hidrógeno y los contactos hidrofóbicos sigue la implementación de la función *pro\_Contacts* en MOE. En este sentido, la implementación realizada difiere de la original, al no considerar *pro\_Contacts* exactamente las

mismas definiciones para los átomos como posibles dadores y aceptores. En principio, se parametriza internamente el valor del radio de definición del sitio activo (10 Å).

## 6.2. Sistemas de *scoring* considerados en el cribado virtual

Una vez calculados los *fingerprints* para la base de datos *dockeada* y una referencia cristalográfica, se utilizan dos sistemas de *scoring* para evaluar la mejor conformación de cada ligando y ordenar los compuestos.

- Los coeficientes de similitud tradicionales (apartado 1.10.1): coeficiente de Tanimoto, distancia Euclídea, distancia Manhattan y el *Simple matching*. Este último corresponde a la suma de los bits puestos a uno y comunes para los dos *fingerprints* a ser comparados, es decir, el numerador del coeficiente de Tanimoto. En adelante se referirá a este sistema de *scoring* como SCORE1, especificando en cada caso el coeficiente utilizado.
- Siguiendo la idea de Mpamhanga<sup>445</sup>, se calcula una segunda serie de *scores* resultantes de la multiplicación de la función de *scoring* del *docking* y del *scoring* de similitud. Para los coeficientes de asociación (Tanimoto y *Simple matching*), dado que un mayor valor indica mayor similitud, la multiplicación es directa. Para los coeficientes de distancia (Euclídea y Manhattan), dado que a mayor similitud se obtiene un menor coeficiente, la multiplicación no es directa y se debería invertir la tendencia (restando el valor de similitud a un valor máximo). Por simplicidad, no se han considerado estos últimos coeficientes en este segundo esquema de *scoring*. En adelante se refiere a este sistema como SCORE2, especificando en cada caso el coeficiente utilizado (Tanimoto o *Simple matching*).

Los resultados se comparan con los obtenidos utilizando el criterio único de la función de *scoring* del *docking*.

## 6.3. Aplicación al cribado virtual de antagonistas del receptor $\alpha$ de estrógeno

Se analiza el comportamiento de estos *fingerprints* en el mismo caso de estudio utilizado por Mpamhanga<sup>445</sup>, en el *data set* de inhibidores del receptor  $\alpha$  de estrógeno (ER $\alpha$ ). El principal objetivo en este caso es constatar si la implementación del *fingerprint* CHIF en este trabajo se adecua a la descrita en la referencia original.

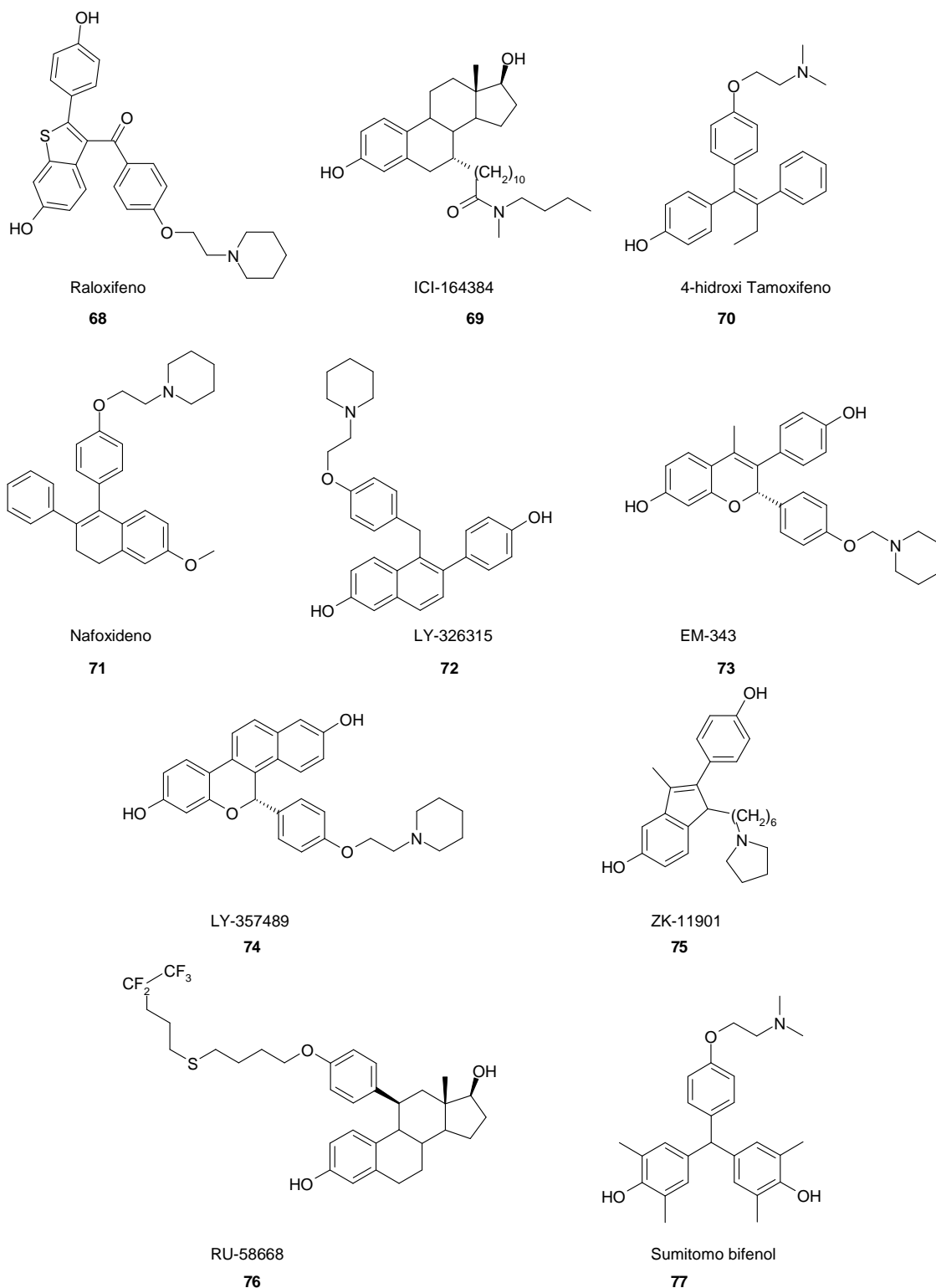
Siguiendo el procedimiento descrito<sup>445</sup>, se considera como estructura de referencia el complejo de ER $\alpha$  con el 4-hidroxi-tamoxifeno (4HT), correspondiente a la entrada PDB 3ert (resolución 1.9 Å).<sup>447</sup> Se eliminan las aguas y se genera directamente el *fingerprint* de referencia sobre este complejo.

La base de datos del cribado virtual se compone de 10 moléculas activas (*"Bissantz active set"*<sup>42</sup>) y 490 moléculas seleccionadas aleatoriamente del catálogo Maybridge, todas ellas asumidas como inactivas y con propiedades 1D focalizadas hacia los activos. Este *pool* de inactivos difiere del utilizado por Mpamhanga (también de 490 compuestos), aunque cabe destacar que estos autores realizaron los cálculos para tres *pools* de inactivos sin encontrar diferencias notables entre ellos, por lo que *a priori* no es esperable que la modificación del *pool* de inactivos trascienda. Las estructuras de los 10 compuestos activos se muestran en la Figura 6.5. Las propiedades 1D de ambos *pools* se listan en la Tabla 6.1.

**Tabla 6.1.** Propiedades 1D del *pool* de activos e inactivos de la base de datos.

	Peso Molecular	b_1rotN	a_acc	a_don
10 Activos	458.8 (67.1)	11.3 (4.3)	3.6 (0.8)	1.6 (0.7)
490 Inactivos	465.1 (16.2)	8.5 (3.1)	4.2 (1.8)	1.0 (0.9)

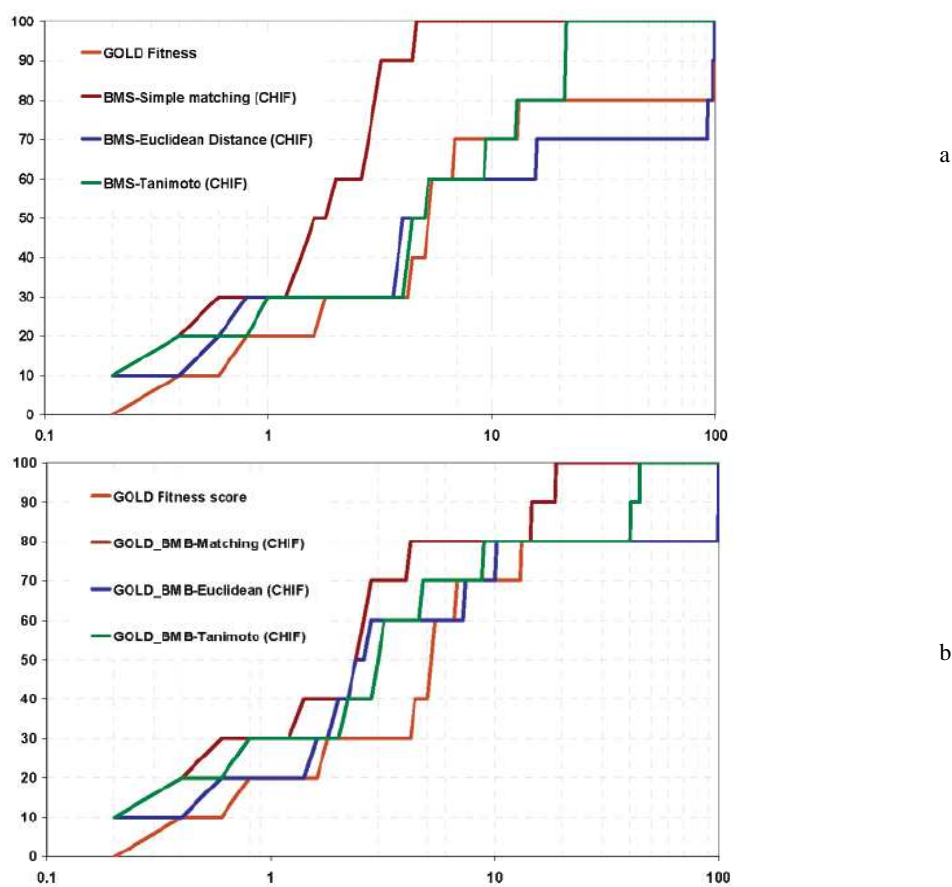
*b\_1rotN*: número de enlaces rotables, *a\_acc*: número de aceptores, *a\_don*: número de dadores.



**Figura 6.5.** Antagonistas de ER $\alpha$  que conforman el *pool* de activos. El 4-hidroxi tamoxifeno es la estructura de referencia co-cristalizada en el sitio de unión (entrada PDB 3ert).

Se realiza el *docking* de estos 500 compuestos frente a 3ert (tras añadir y relajar la posición de los hidrógenos) con el programa GOLD 3.0.1 y la función de *scoring* GOLDScore. Para cada compuesto se requieren 10 *dockings* independientes (conórmeros), con un valor máximo de 100.000 operaciones genéticas y un tamaño de población del algoritmo genético de 100 cromosomas. El resto de parámetros de GOLD se mantiene por defecto. La cavidad de unión se define a partir de un residuo situado en la cavidad de unión (Leu<sup>525</sup>) y con un radio de 10 Å, activándose el algoritmo de detección de cavidad.

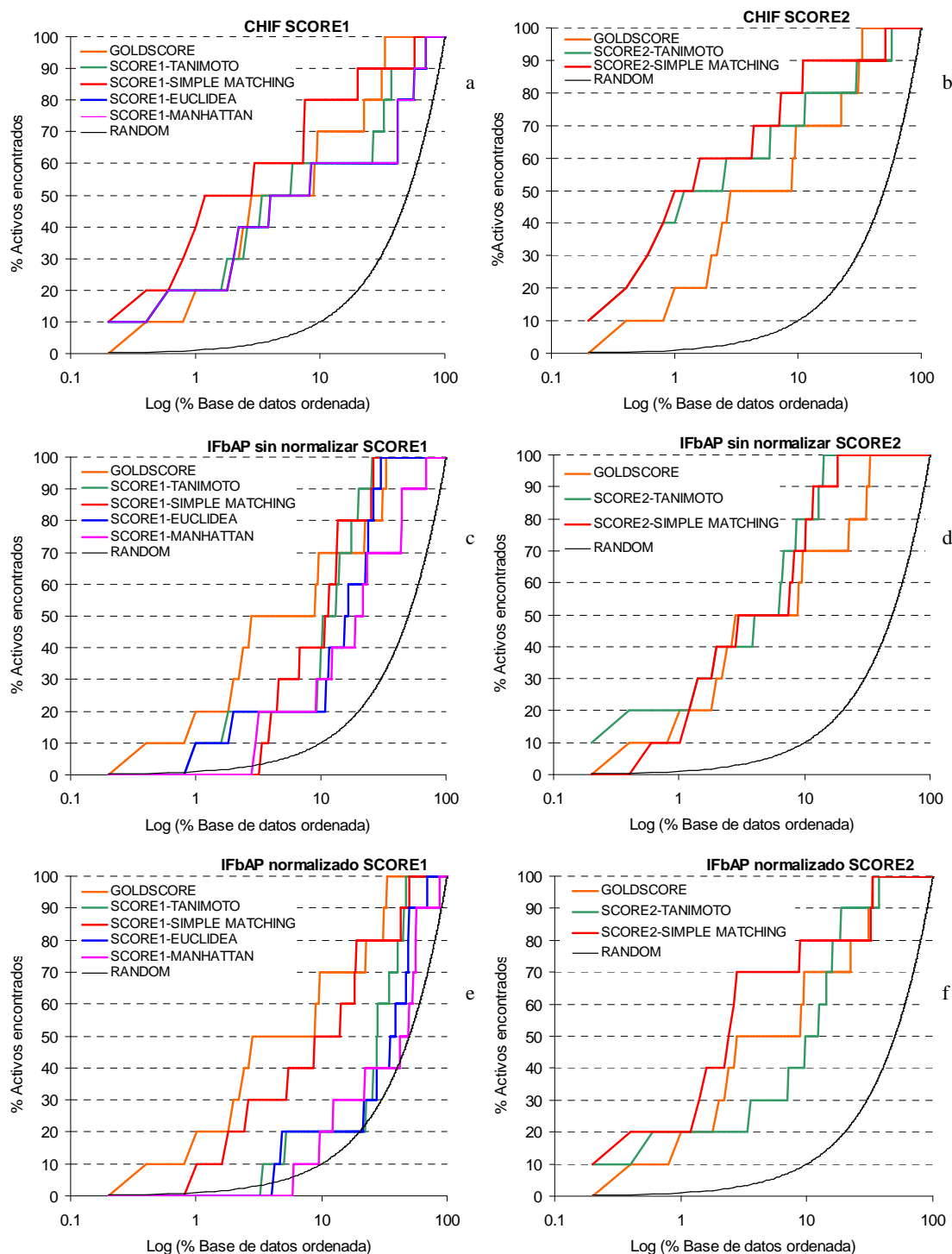
En la Figura 6.6 se muestran las curvas de enriquecimiento presentadas por Mpamhanga en el cribado virtual con el *fingerprint* CHIF, ordenando los compuestos por similitud decreciente al *fingerprint* de referencia (SCORE1, Figura 6.6.a) u ordenándolos por el producto de su similitud y el valor de GOLDScore (SCORE2, Figura 6.6.b). Para realzar la parte inicial de la curva, se representa en el eje de abscisas el porcentaje de base de datos cribada en escala logarítmica.



**Figura 6.6.** Curvas de enriquecimiento presentadas por Mpamhanga utilizando el *fingerprint* CHIF y ordenando los compuestos según (a) el *scoring* de similitud, SCORE1 y (b) el producto del *scoring* de similitud y el valor de la función de *fitness* de GOLDScore, SCORE2. Los diferentes coeficientes de similitud utilizados corresponden a *Simple matching* (rojo), distancia Euclídea (azul) y coeficiente de Tanimoto (verde). La curva de enriquecimiento obtenida con GOLDScore se muestra en naranja. El eje de abscisas es la escala logarítmica del porcentaje de base de datos cribada y el eje de ordenadas corresponde al porcentaje de activos encontrados. Extraído de [445].

En la Figura 6.7 se muestran las curvas de enriquecimiento obtenidas con la implementación del *fingerprint* CHIF en este trabajo (Figuras 6.7.a y 6.7.b). Se observa cómo no existen diferencias significativas con los resultados mostrados en la Figura 6.6, por lo que se da por validada la implementación realizada, a pesar de las diferencias en la definición de puentes de hidrógeno y contactos hidrofóbicos. Se muestran también en esta figura, las curvas de enriquecimiento

obtenidas con la implementación de IFbAP sin normalizar (Figuras 6.7.c y 6.7.d) y normalizado entre 0 y 1 (Figuras 6.7.e y 6.7.f).



**Figura 6.7.** Curvas de enriquecimiento obtenidas en el cribado de inhibidores de ER $\alpha$  con la implementación en MOE de (a) CHIF y *scoring* de similitud (SCORE1). (b) CHIF y el producto de *scoring* de similitud y el valor de la función de *fitness* de GOLDScore (SCORE2). (c) IFbAP sin normalizar y SCORE1. (d) IFbAP sin normalizar y SCORE2. (e) IFbAP normalizado y SCORE1. (f) IFbAP normalizado y SCORE2.

Los diferentes coeficientes de similitud utilizados corresponden a *Simple matching* (rojo), distancia Euclídea (azul), coeficiente de Tanimoto (verde) y distancia Manhattan (rosa). Cribado virtual con GOLDScore (naranja) y cribado aleatorio (negro).

Aunque IFbAP permite recuperar los compuestos por encima de una selección aleatoria, el enriquecimiento generado considerando únicamente la similitud al *fingerprint* de referencia se sitúa por debajo del enriquecimiento obtenido con la función de *scoring* GOLDScore (Figuras 6.7.c y 6.7.e). La combinación de los criterios de similitud y energético (SCORE2) alcanza rendimientos superiores a los obtenidos únicamente con el criterio energético (Figuras 6.7.d y 6.7.f). Sin embargo, aún considerando SCORE2, IFbAP no alcanza el elevado rendimiento de CHIF en el primer 1-2% de base de datos cribada, aunque sí a porcentajes superiores.

Respecto a la normalización, no se observa una tendencia definitiva. Mientras que para SCORE1 la normalización empeora el comportamiento (Figura 6.7.c *versus* Figura 6.7.e), para SCORE2 tiene un efecto positivo (Figura 6.7.d *versus* Figura 6.7.f).

En la Tabla 6.2 se muestran los factores de enriquecimiento en el primer 2%, 5% y 10% de base de datos cribada. El valor de *ef* máximo es de 50 (500/10). En negrita se muestran, para cada porcentaje, el valor máximo encontrado. Se observa cómo CHIF con SCORE2 es el óptimo claramente y cómo IFbAP con SCORE1 no discrimina tan bien entre activos e inactivos como lo hace el criterio energético del *docking* (GOLDScore), aunque su comportamiento mejora al combinarse con el criterio energético del *docking* (IFbAP\_NORMALIZADO-SCORE2-SIMPLE\_MATCHING *versus* GOLDScore).

**Tabla 6.2.** Factores de enriquecimiento (*ef*) en el primer 2%, 5% y 10% de base de datos cribada.

	2%	5%	10%
GOLDScore ( <i>docking</i> )	15	10	7
CHIF-SCORE1-TANIMOTO	15	10	6
CHIF-SCORE1-SIMPLE_MATCHING	25	12	<b>8</b>
CHIF-SCORE1-EUCLÍDEA	15	10	6
CHIF-SCORE1-MANHATTAN	15	10	6
CHIF-SCORE2-TANIMOTO	25	12	7
CHIF-SCORE2-SIMPLE_MATCHING	<b>30</b>	<b>14</b>	<b>8</b>
IFbAP-SCORE1-TANIMOTO	10	4	4
IFbAP-SCORE1-SIMPLE_MATCHING	0	6	4
IFbAP-SCORE1-EUCLÍDEA	10	4	2
IFbAP-SCORE1-MANHATTAN	0	4	3
IFbAP-SCORE2-TANIMOTO	20	10	<b>8</b>
IFbAP-SCORE2-SIMPLE_MATCHING	20	10	7
IFbAP_NORMALIZADO-SCORE1-TANIMOTO	0	2	2
IFbAP_NORMALIZADO-SCORE1-SIMPLE_MAT	10	6	5
IFbAP_NORMALIZADO-SCORE1-EUCLÍDEA	0	4	2
IFbAP_NORMALIZADO-SCORE1-MANHATTAN	0	0	2
IFbAP_NORMALIZADO-SCORE2-TANIMOTO	10	6	5
IFbAP_NORMALIZADO-SCORE2-SIMPLE_MAT	20	<b>14</b>	<b>8</b>

#### 6.4. Aplicación al cribado virtual de inhibidores de FGFR

Se analiza la bondad de los dos *fingerprints* (CHIF e IFbAP) en el cribado virtual frente a tirosina quinasas. Se escoge el cribado virtual por *docking* con AUTODOCK de la base de datos Base\_Maybridge frente a la estructura 2fgi de FGFR (apartado 5.2.4). Base\_Maybridge contiene 288 inhibidores de FGFR y 4342 presuntos inactivos extraídos del catálogo Maybridge.

Tanto para CHIF como para IFbAP, se adopta como *fingerprint* de referencia el calculado para el complejo de FGFR con PD173074 (entrada PDB 2fgi).

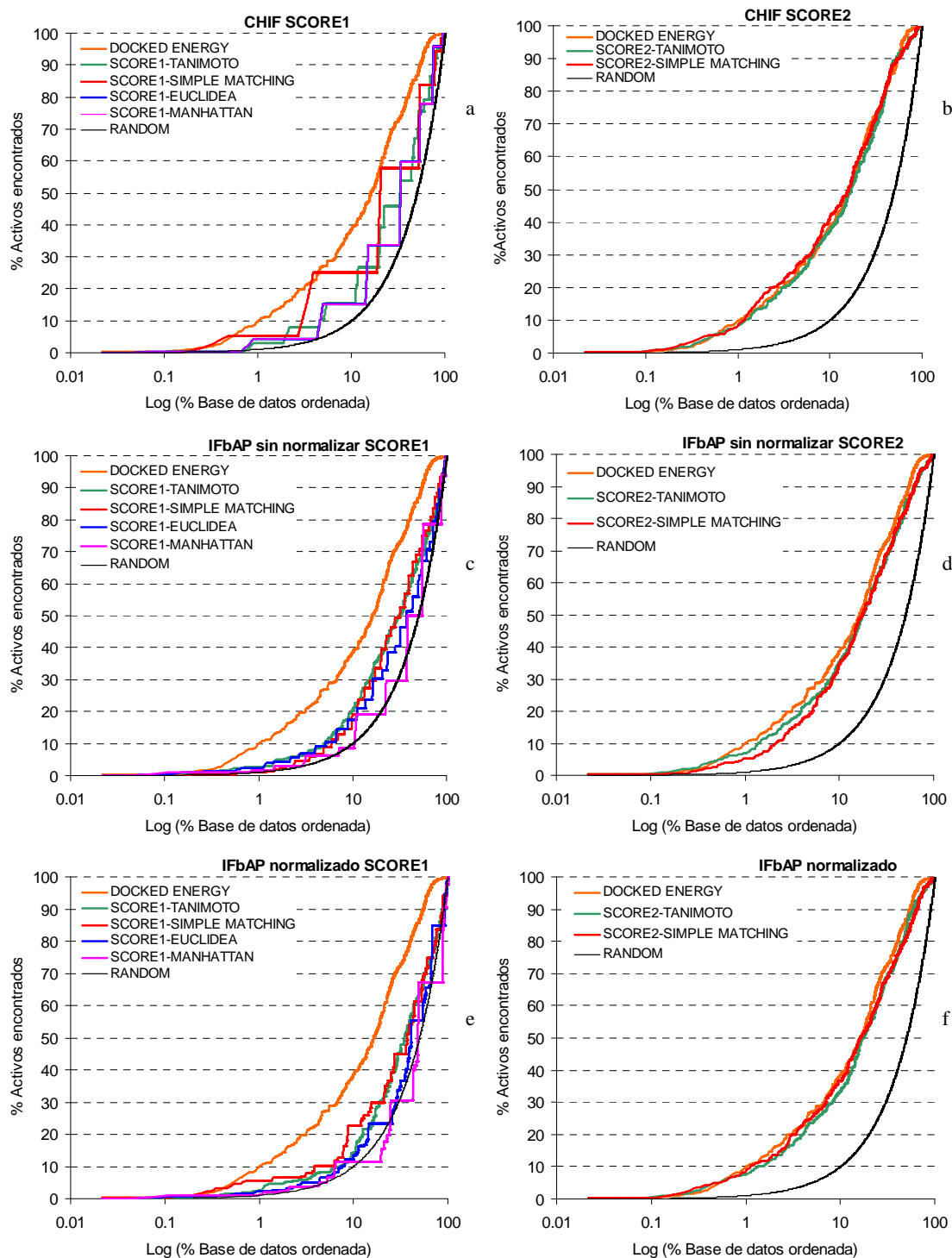
En la Figura 6.8 se muestran las curvas de enriquecimiento obtenidas para los dos *fingerprints* y bajos los dos sistemas de *scoring* propuestos. En este caso particular, CHIF con SCORE1 no recupera compuestos activos a una tasa superior a la del criterio energético de AUTODOCK (*docked energy*), aunque se sitúa por encima de una selección aleatoria (Figura 6.8.a). Combinar las medidas de similitud de CHIF con la función de *scoring* del *docking* equivale prácticamente a la discriminación que efectúa el criterio energético por sí solo (Figura 6.8.b).

El mismo comportamiento se observa con el *fingerprint* IFbAP, tanto normalizado como sin normalizar. De hecho, las medidas de similitud establecidas con este *fingerprint* establecen un ranking de la base de datos que discrimina peor que el método CHIF (Figuras 6.8.c y 6.8.e *versus* 6.8.a). Esto se puede apreciar en los valores de los factores de enriquecimiento mostrados en la Tabla 6.3 para el primer 2%, 5% y 10% de base de datos cribada. En este caso, el valor de *ef* máximo teórico es de 16.1 (4630/288).

**Tabla 6.3.** Factores de enriquecimiento (*ef*) en el primer 2%, 5% y 10% de base de datos cribada.

	2%	5%	10%
GOLDScore ( <i>docking</i> )	8.0	<b>5.4</b>	3.9
CHIF-SCORE1-TANIMOTO	2.6	2.1	1.6
CHIF-SCORE1-SIMPLE_MATCHING	2.6	5.0	2.5
CHIF-SCORE1-EUCLÍDEA	2.1	3.0	1.5
CHIF-SCORE1-MANHATTAN	2.1	3.0	1.5
CHIF-SCORE2-TANIMOTO	6.9	5.0	3.7
CHIF-SCORE2-SIMPLE_MATCHING	<b>9.2</b>	5.0	<b>4.2</b>
IFbAP-SCORE1-TANIMOTO	2.6	2.0	2.0
IFbAP-SCORE1-SIMPLE_MATCHING	1.6	1.8	1.9
IFbAP-SCORE1-EUCLÍDEA	2.1	1.9	1.7
IFbAP-SCORE1-MANHATTAN	1.6	1.2	0.9
IFbAP-SCORE2-TANIMOTO	6.4	4.7	3.5
IFbAP-SCORE2-SIMPLE_MATCHING	4.5	3.9	3.5
IFbAP_NORMALIZADO-SCORE1-TANIMOTO	2.6	1.7	1.5
IFbAP_NORMALIZADO-SCORE1-SIMPLE_MAT	3.3	2.0	2.3
IFbAP_NORMALIZADO-SCORE1-EUCLÍDEA	1.6	1.3	1.3
IFbAP_NORMALIZADO-SCORE1-MANHATTAN	1.4	1.0	1.1
IFbAP_NORMALIZADO-SCORE2-TANIMOTO	6.2	4.7	3.0
IFbAP_NORMALIZADO-SCORE2-SIMPLE_MAT	6.9	<b>5.1</b>	3.6





**Figura 6.8** Curvas de enriquecimiento obtenidas en el **cribado de inhibidores de FGFR** con la implementación en MOE de (a) CHIF y *scoring* de similitud (SCORE1). (b) CHIF y el producto de *scoring* de similitud y el valor de la función de *docked energy* de AUTODOCK (SCORE2). (c) IFbAP sin normalizar y SCORE1. (d) IFbAP sin normalizar y SCORE2. (e) IFbAP normalizado y SCORE1. (f) IFbAP normalizado y SCORE2.

Los diferentes coeficientes de similitud utilizados corresponden a *Simple matching* (rojo), distancia Euclídea (azul), coeficiente de Tanimoto (verde) y distancia Manhattan (rosa). Cribado virtual con la función *docked energy* de AUTODOCK (naranja) y cribado aleatorio (negro).

## 6.5 Aplicación al cribado virtual de inhibidores de la transcriptasa reversa

El tercer caso de estudio donde se aplican estos *fingerprints* es para el cribado virtual retrospectivo de inhibidores de la transcriptasa reversa (RT) del HIV-1. En los últimos años se han identificado un gran número de compuestos que inhiben el sitio alostérico de RT, denominados *non-nucleoside reverse transcriptase inhibitors* (NNRTIs).

El *pool* de activos contiene 18 NNRTIs, extraídos de la referencia [220], cuyas estructuras se muestran en la Figura 6.9. El *pool* de inactivos se compone de los mismos 490 compuestos incluidos en el cribado de inhibidores de ER $\alpha$  (apartado 6.3). Aunque no se realiza un esfuerzo adicional por focalizar las propiedades 1D del *pool* de inactivos hacia las de los activos (Tabla 6.4), cabe destacar que las estructuras en principio favorecidas por el *docking* lo serían en este caso las inactivas (las de mayor peso molecular).

**Tabla 6.4.** Propiedades 1D del *pool* de activos e inactivos de la base de datos.

	Peso Molecular	b_1rotN	a_acc	a_don
18 Activos	323.3 (43.6)	6.7 (3.1)	2.6 (1.2)	1.1 (0.7)
490 Inactivos	465.1 (16.2)	8.5 (3.1)	4.2 (1.8)	1.0 (0.9)

*b\_1rotN*: número de enlaces rotables, *a\_acc*: número de aceptores, *a\_don*: número de dadores.

Como estructura de referencia se considera el complejo de RT con el inhibidor 8-cloro TIBO (**87**), correspondiente a la entrada PDB 1uwb (resolución 3.2 Å)<sup>448</sup> (Figura 6.3).

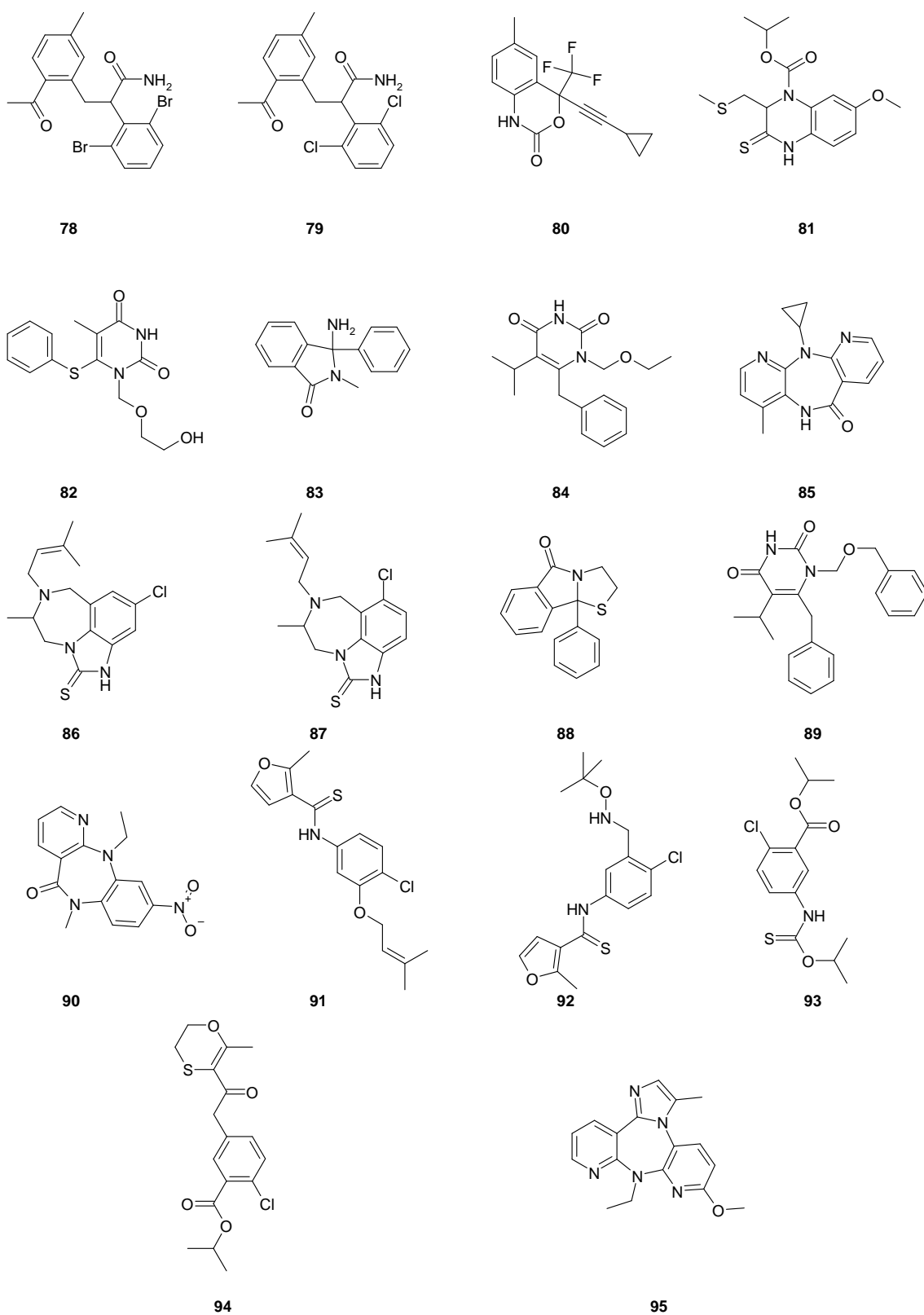
Se acondiciona la estructura de RT para el *docking* de la base de datos con el programa GOLD, manteniéndose las dos subunidades que componen la RT (p66 y p51) y eliminando el 8-cloro TIBO. Se añaden los hidrógenos y se relajan sus posiciones (apartado 5.1.1) con el programa MOE.

Se realiza el *docking* de estos 508 compuestos frente a 1uwb con el programa GOLD 3.0.1 y la función de *scoring* GOLDScore. La cavidad de unión se define a partir de las coordenadas del carbonilo de Lys<sup>101</sup>, con el que 8-cloro TIBO establece un puente de hidrógeno, con un radio de 10 Å. Los restantes parámetros son similares a los descritos en el apartado 6.3.

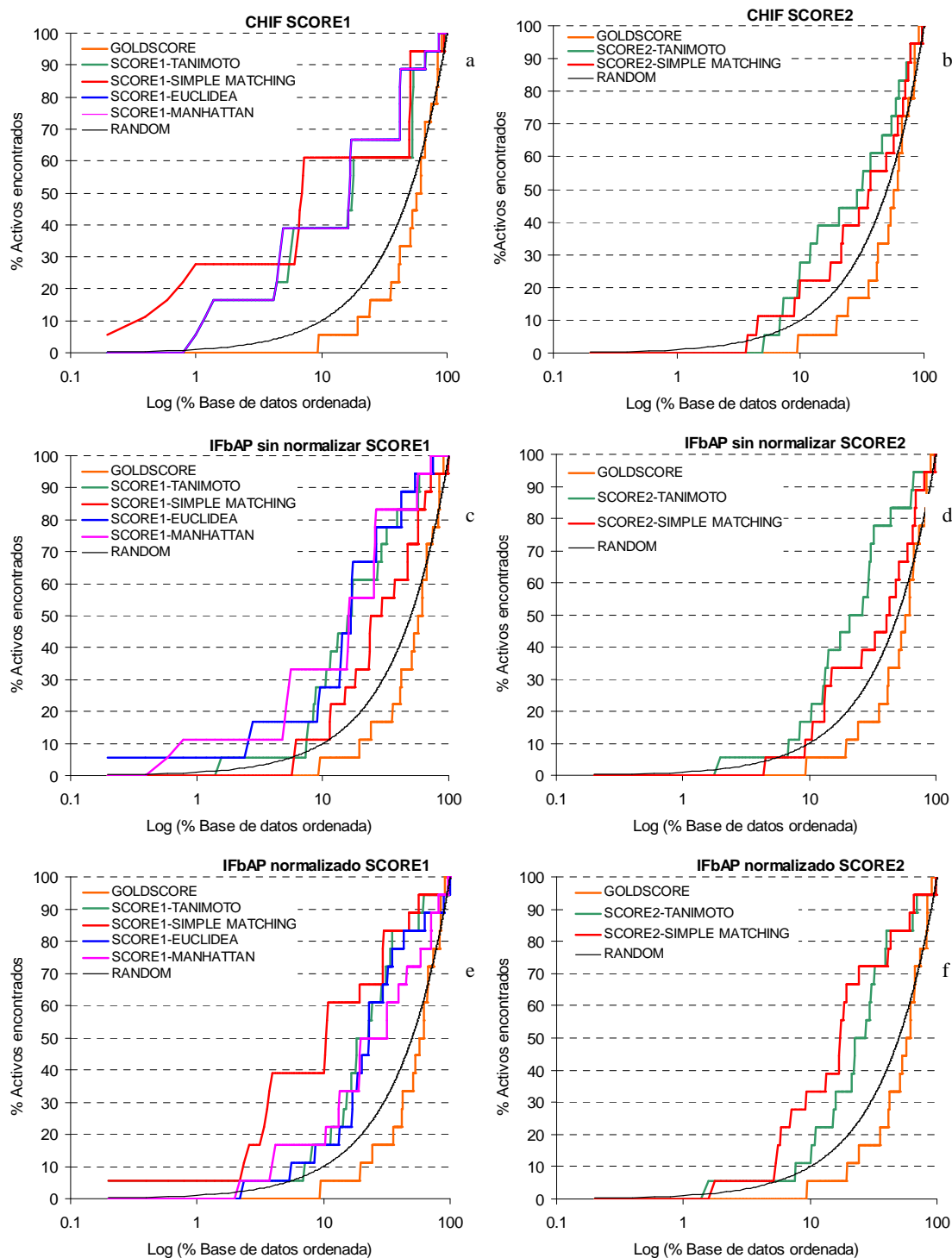
Tras el *docking*, se calculan los *fingerprints* CHIF e IFbAP para todas las conformaciones de todos los compuestos. En la Figura 6.10 se muestran las curvas de enriquecimiento al ordenar los compuestos según su similitud al *fingerprint* de referencia (SCORE1) y según la combinación de criterios de similitud y energético (SCORE2). En la Tabla 6.5 se muestran los factores de enriquecimiento obtenidos en el primer 2%, 5% y 10% de base de datos cribada. El enriquecimiento máximo teórico es de 28.2 (508/18).

En este caso, el *docking* por sí solo no funciona nada bien, quedando por detrás de una selección aleatoria (curva naranja en las gráficas). La búsqueda de similitud con los *fingerprints* (SCORE1) permite discriminar satisfactoriamente entre compuestos activos e inactivos, tanto con CHIF (Figura 6.10.a) como con IFbAP normalizado (Figura 6.10.c) o sin normalizar (Figura 6.10.e). Sin embargo, de nuevo los resultados son mejores con CHIF que con el método propuesto (IFbAP). En los tres casos, obviamente, al incluir el criterio energético en el *scoring* (SCORE2), disminuyen los enriquecimientos (Figuras 6.10.b, 6.10.d y 6.10.f).

Considerando la normalización, se constata de nuevo la falta de consenso. Por ejemplo, con IFbAP sin normalizar y SCORE1 (Figura 6.10.c), los coeficientes que mejor se comportan son la distancia Manhattan y Euclídea. Al normalizar (Figura 6.10.e), el coeficiente *Simple matching* los supera.



**Figura 6.9.** Inhibidores NNRTIs de la transcriptasa reversa que conforman el *pool* de activos. El TIBO (87) es la estructura de referencia co-cristalizada en el sitio de unión (entrada PDB 1uwb).



**Figura 6.10** Curvas de enriquecimiento obtenidas en el **cribado de inhibidores de RT** con la implementación en MOE de (a) CHIF y *scoring* de similitud (SCORE1). (b) CHIF y el producto de *scoring* de similitud y GOLDScore (SCORE2). (c) IFbAP sin normalizar y SCORE1. (d) IFbAP sin normalizar y SCORE2. (e) IFbAP normalizado y SCORE1. (f) IFbAP normalizado y SCORE2. Los diferentes coeficientes de similitud utilizados corresponden a *Simple matching* (rojo), distancia Euclídea (azul), coeficiente de Tanimoto (verde) y distancia Manhattan (rosa). Cribado virtual con la función GOLDScore (naranja) y cribado aleatorio (negro).

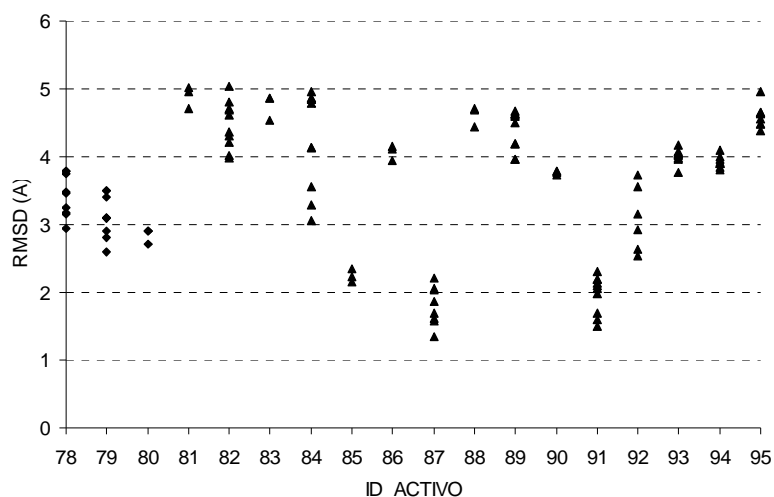
**Tabla 6.5.** Factores de enriquecimiento (*ef*) en el primer 2%, 5% y 10% de base de datos cribada.

	2%	5%	10%
GOLDScore ( <i>docking</i> )	0.0	0.0	0.6
CHIF-SCORE1-TANIMOTO	7.7	4.3	3.9
CHIF-SCORE1-SIMPLE_MATCHING	12.8	5.4	6.1
CHIF-SCORE1-EUCLÍDEA	7.7	7.6	3.9
CHIF-SCORE1-MANHATTAN	7.7	7.6	3.9
CHIF-SCORE2-TANIMOTO	0.0	1.1	2.8
CHIF-SCORE2-SIMPLE_MATCHING	0.0	2.2	2.2
IFbAP-SCORE1-TANIMOTO	2.6	1.1	2.8
IFbAP-SCORE1-SIMPLE_MATCHING	0.0	0.0	1.1
IFbAP-SCORE1-EUCLÍDEA	2.6	3.3	2.8
IFbAP-SCORE1-MANHATTAN	5.1	4.3	3.3
IFbAP-SCORE2-TANIMOTO	2.6	1.1	1.7
IFbAP-SCORE2-SIMPLE_MATCHING	0.0	1.1	1.1
IFbAP_NORMALIZADO-SCORE1-TANIMOTO	2.6	1.1	1.7
IFbAP_NORMALIZADO-SCORE1-SIMPLE_MAT	2.6	7.6	3.9
IFbAP_NORMALIZADO-SCORE1-EUCLÍDEA	0.0	1.1	1.7
IFbAP_NORMALIZADO-SCORE1-MANHATTAN	2.6	3.3	1.7
IFbAP_NORMALIZADO-SCORE2-TANIMOTO	2.6	1.1	1.1
IFbAP_NORMALIZADO-SCORE2-SIMPLE_MAT	2.6	1.1	3.3

Existen estructuras resueltas de 14 activos co-cristalizados con RT y 4 modelos teóricos depositados en el *Protein Data Bank* (tabulados en la Tabla 6.6) por lo que es factible en este caso realizar una medida de la RMSD entre las conformaciones resultantes del *docking* y el ligando correspondiente. Estos resultados se resumen en la Figura 6.11. Se observa como excepto para dos compuestos, **87** y **91** que se sitúan por debajo de 2 Å, para el resto de ligandos las conformaciones se encuentran mayoritariamente entre 3 y 4 Å. A la vista de la mala predicción del modo de unión, no sorprende que el cribado virtual con GOLDScore sea tan deficiente. El hecho de que los *fingerprints* funcionen tan bien en este caso indica que se conservan las interacciones principales, pese a que las conformaciones no están perfectamente orientadas (principalmente, considerando los resultados de CHIF, procedimiento validado externamente).

**Tabla 6.6.** Lista de inhibidores NNRTIs co-cristalizados con RT y sus entradas PDB.

Compuesto	Entrada PDB	Técnica/resolución	Referencia
78	1hni	Rayos-X. 2.80 Å	449
79	1vru	Rayos-X. 2.40 Å	450
80	1fk9	Rayos-X. 2.50 Å	451
81	1bqm	Rayos-X. 3.10 Å	452
82	1rvm	Modelo teórico	453
83	1rvn	Modelo teórico	453
84	1rt1	Rayos-X. 2.55 Å	454
85	1vrt	Rayos-X. 2.20 Å	450
86	1tvr	Rayos-X. 3.00 Å	448
87	1uwb	Rayos-X. 3.20 Å	448
88	1rvp	Modelo teórico	453
89	1rt2	Rayos-X. 2.55 Å	454
90	1rt3	Rayos-X. 3.00 Å	455
91	1rt4	Rayos-X. 2.90 Å	456
92	1rt5	Rayos-X. 2.90 Å	456
93	1rt6	Rayos-X. 2.80 Å	456
94	1rt7	Rayos-X. 3.00 Å	456
95	1rvr	Modelo teórico	453

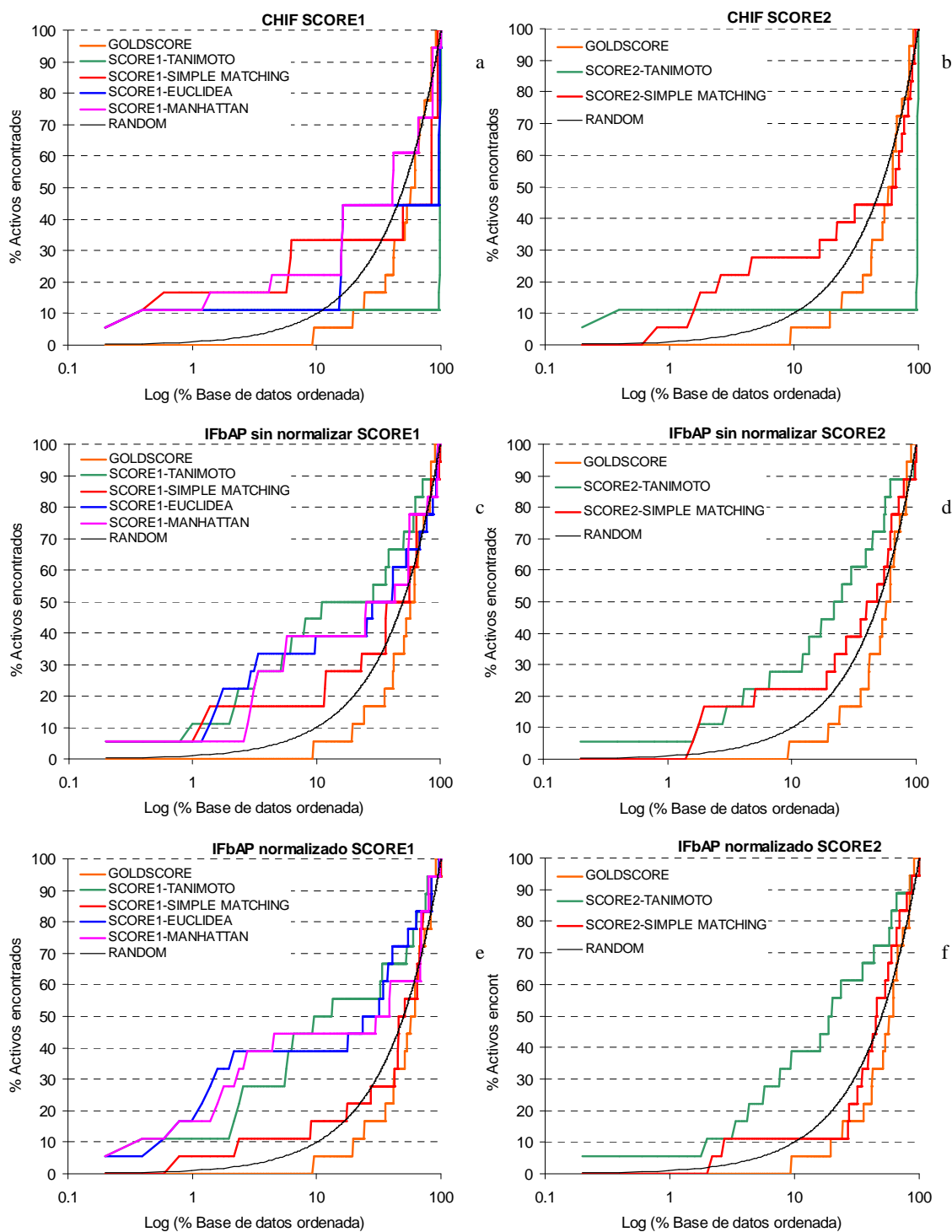


**Figura 6.11.** RMSD de superposición de todos los conformeros obtenidos para cada uno de los 18 inhibidores de RT (Figura 6.8)

La mala predicción del modo de unión por el *docking* afecta tanto al ranking de los compuestos por el criterio energético (GOLDScore) como al comportamiento de los *fingerprints*. Así, se decide repetir los cálculos para las conformaciones de los activos correctamente orientadas en el sitio de unión. Para ello, se superponen los complejos cristalográficos (Tabla 6.6) frente al complejo 1uwb, considerando únicamente los residuos alrededor del sitio alostérico de unión: 92-110, 227-236, 316-319 de la cadena A y 138 de la cadena B. Se transfieren las conformaciones de los ligandos al sitio de unión de 1uwb. Las conformaciones de los inactivos corresponden a las obtenidas en el *docking* con GOLDScore. Para las conformaciones de los activos se recalcula el valor de *scoring* con GOLDScore y se recalculan los nuevos *fingerprints* (CHIF e IFbAP). En la Figura 6.12 se muestran las curvas de enriquecimiento obtenidas y en la Tabla 6.7 los nuevos factores de enriquecimiento.

**Tabla 6.7.** Factores de enriquecimiento (*ef*) en el primer 2%, 5% y 10% de base de datos cribada.

	2%	5%	10%
GOLDScore ( <i>docking</i> )	0.0	0.0	0.8
CHIF-SCORE1-TANIMOTO	5.1	2.2	1.1
CHIF-SCORE1-SIMPLE_MATCHING	7.7	3.3	3.3
CHIF-SCORE1-EUCLÍDEA	5.1	2.2	1.1
CHIF-SCORE1-MANHATTAN	7.7	4.3	2.2
CHIF-SCORE2-TANIMOTO	5.1	2.2	1.1
CHIF-SCORE2-SIMPLE_MATCHING	7.7	5.4	2.8
IFbAP-SCORE1-TANIMOTO	7.7	5.4	4.4
IFbAP-SCORE1-SIMPLE_MATCHING	7.7	3.3	1.7
IFbAP-SCORE1-EUCLÍDEA	10.3	6.5	3.9
IFbAP-SCORE1-MANHATTAN	2.6	5.4	3.9
IFbAP-SCORE2-TANIMOTO	5.1	4.3	2.8
IFbAP-SCORE2-SIMPLE_MATCHING	7.7	4.3	2.2
IFbAP_NORMALIZADO-SCORE1-TANIMOTO	7.7	5.4	5.0
IFbAP_NORMALIZADO-SCORE1-SIMPLE_MAT	2.6	2.2	1.7
IFbAP_NORMALIZADO-SCORE1-EUCLÍDEA	18.0	7.6	3.9
IFbAP_NORMALIZADO-SCORE1-MANHATTAN	12.8	8.7	4.4
IFbAP_NORMALIZADO-SCORE2-TANIMOTO	5.1	4.3	3.9
IFbAP_NORMALIZADO-SCORE2-SIMPLE_MAT	2.6	2.2	1.1



**Figura 6.12.** Curvas de enriquecimiento obtenidas en el **cribado de inhibidores de RT superpuestos** con la implementación en MOE de (a) CHIF y *scoring* de similitud (SCORE1). (b) CHIF y el producto de *scoring* de similitud y GOLDScore (SCORE2). (c) IFbAP sin normalizar y SCORE1. (d) IFbAP sin normalizar y SCORE2. (e) IFbAP normalizado y SCORE1. (f) IFbAP normalizado y SCORE2. Las conformaciones de los activos corresponden a las obtenidas por superposición de los correspondientes complejos cristalográficos frente a luwb. Los diferentes coeficientes de similitud utilizados corresponden a *Simple matching* (rojo), distancia Euclídea (azul), coeficiente de Tanimoto (verde) y distancia Manhattan (rosa). Cribado virtual con la función GOLDScore (naranja) y cribado aleatorio (negro).

Se observa cómo en este caso, de nuevo el ranking establecido con la función GOLDScore opera mal, esta vez atribuyéndose la causa a la función de *scoring* y no a la predicción del modo de unión. Sorprende el hecho de que el *fingerprint* CHIF empeore su comportamiento respecto al caso anterior (Figura 6.12.a *versus* Figura 6.10.a), donde las *poses* proceden de un *docking* con una RMSD de 3-4 Å respecto a la conformación cristalográfica. Por otra parte, el comportamiento de IFbAP mejora considerablemente, tanto normalizado como sin normalizar (Figura 6.12.c *versus* Figura 6.10.c y Figura 6.12.e *versus* Figura 6.10.e). Este comportamiento sería el esperable, ya que en este caso las conformaciones corresponden a las obtenidas por superposición de las cristalográficas.

En conclusión, el *fingerprint* IFbAP desarrollado, si bien discrimina compuestos activos de inactivos por encima de una selección aleatoria, no se comporta tan bien como el *fingerprint* CHIF para la mayor parte de los casos, excepto para la transcriptasa reversa con las conformaciones de los activos superpuestas en la cavidad de unión. Para ambos, la combinación del criterio de similitud con el criterio energético permite incrementar el enriquecimiento, excepto si las conformaciones del *docking* son deficientes o si la función de *scoring* se comporta mal, como sucede en el cribado frente a la transcriptasa reversa.

Parece que ambos métodos (CHIF e IFbAP) operan mejor en casos donde las interacciones en el complejo son principalmente de tipo hidrofóbico (caso del receptor  $\alpha$  de estrógeno) que en casos donde las interacciones básicas entre ligando y receptor son por puente de hidrógeno (caso de FGFR). Esto puede asumirse principalmente en la descripción IFbAP, donde el mayor número de interacciones hidrofóbicas adquiere más importancia que las interacciones por puente de hidrógeno (véase Figura 6.4). Así, con ánimo de mejorar IFbAP, se podría plantear la introducción de algún tipo de ponderación o escalado que permita incrementar el peso de las minoritarias interacciones por puente de hidrógeno.



## Capítulo 7.

### Predicción de la Afinidad de Unión

En este capítulo se estudia la predicción de la afinidad de unión para una serie de inhibidores de FGFR (apartado 4.1). Como métodos de cálculo de energía libre de unión se consideran tanto las funciones empíricas implementadas en los dos programas de *docking* utilizados (AUTODOCK y GOLD), como el método MM-PBSA/MM-GBSA descrito en el apartado 1.4.2. Posteriormente, este método se aplica en el cribado virtual de una base de datos frente a FGFR.

#### 7.1. Predicción de afinidad frente FGFR

La relación entre  $IC_{50}$  experimental (concentración de ligando necesaria para inhibir un 50% la actividad *in vitro*) y la energía libre de unión se establece a partir de la ecuación [7.1]. Mientras que  $K_i$  (constante de inhibición) y  $\Delta G$  sí que presentan una relación estrictamente lineal, la correlación entre  $IC_{50}$  y  $\Delta G$  no es exactamente lineal.

$$\Delta G \approx RT \ln(IC_{50}) \quad [7.1]$$

Así, más que predecir un valor absoluto de energía de unión, resulta más interesante la predicción del ranking de las moléculas candidatas, es decir, su orden relativo. Existen dos coeficientes que permiten cuantificar el grado de correlación entre el orden de las moléculas en función de su actividad calculada y experimental:

- El coeficiente de correlación de Spearman ( $R_s$ )<sup>457</sup> (ecuación [7.2]) calcula la correlación entre dos sets de rankings:

$$R_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad [7.2]$$

donde  $d_i$  es la diferencia en la posición en el ranking para una molécula candidata bajo dos criterios de ordenación y  $n$  es el número de objetos en el ranking, en este caso, el número de moléculas candidatas. Aquí, uno de los criterios es la afinidad de unión experimental y el otro es la energía de unión calculada por uno de los métodos. En teoría, este coeficiente adopta valores comprendidos entre -1 y +1, donde +1 corresponde a una correlación perfecta, -1 a una correlación inversa y 0 indica ausencia de correlación.

- El índice predictivo (*predictive index*, PI), introducido por Pearlman.<sup>458</sup>

$$PI = \frac{\sum_{j>i} \sum_i w_{ij} C_i}{\sum_{j>i} \sum_i w_{ij}} \quad [7.3]$$

donde:

$$w_{ij} = |E(j) - E(i)| \quad [7.4]$$

$$C_{ij} = \begin{cases} 1 & \text{si } |E(j) - E(i)| / |P(j) - P(i)| < 0 \\ -1 & \text{si } |E(j) - E(i)| / |P(j) - P(i)| > 0 \\ 0 & \text{si } |P(j) - P(i)| = 0 \end{cases} \quad [7.5]$$

$P(i)$  corresponde al valor de *scoring* (energía de unión) asignado a una molécula  $i$  por el método correspondiente y  $E(i)$  es el valor experimental (expresado como  $pIC_{50}$  o  $pK_i$ ). Este índice adopta valores comprendidos entre -1 y +1. Como en el caso de  $R_s$ , +1 corresponde a una correlación perfecta y -1 a predicciones inversas, mientras que 0 indica predicciones completamente aleatorias. A diferencia de  $R_s$ , este coeficiente contiene un término ( $w_{ij}$ ) que pondera la diferencia entre los valores de energía experimentales, reflejando así el hecho de que una función debería ser capaz de discernir mejor entre aquellos cambios estructurales que resulten en una mayor diferencia en la unión. La convención de signos establecida en la ecuación [7.5] debe revisarse en función de si se representa  $IC_{50}$  en lugar de  $pIC_{50}$  o de cómo es la función energética (función de *scoring*), es decir, si comprende un rango de valores positivos o valores negativos.

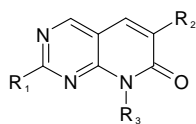
Para realizar el estudio de predicción de afinidad se escogen 27 compuestos de la serie de inhibidores pirido[2,3-*d*]pirimidinas, con actividades comprendidas en un rango de 4 órdenes de magnitud: con  $IC_{50}$  desde 30 nM a 13.9  $\mu$ M ( $pIC_{50}$  de 4.8 a 7.5) frente FGFR. Todos ellos han sido testados en las mismas condiciones experimentales. Las estructuras de estos inhibidores, así como su  $IC_{50}$  ( $\mu$ M) frente a las tres quinasas PDGFR, FGFR y EGFR se muestran en la Tabla 7.1. Su número de identidad corresponde al de las estructuras incluidas en el Anexo 2 y  $R^1$ ,  $R^2$  y  $R^3$  se refieren a los tres puntos de sustitución del *scaffold* mostrado en la Figura 7.1.

Como estructura cristalina del dominio tirosina quinasa de FGFR se escoge 2fgi en base a los resultados de predicción del modo de unión presentados en el apartado 5.1 para estos inhibidores del tipo pirido[2,3-*d*]pirimidinas.

### 7.1.1. Procedimiento para el cálculo de energías libres de unión

Se repite el *docking* de estos 27 compuestos frente a FGFR (2fgi), en condiciones similares a las descritas en el apartado 5.1.2 aunque incrementando el número de conformeros (*dockings* independientes) de 10 a 100. Como funciones de *scoring*/programas de *docking* se escogen:

- AUTODOCK. Se analizan cuatro combinaciones iniciales, resultantes de aplicar dos tamaños de malla (101×101×101 y 61×61×61) y dos sistemas de cargas parciales: Gasteiger-Marsili y AM1. Para la predicción de afinidad se utiliza la evaluación energética *binding free energy*. Se analizan los resultados para el conformero de mínima energía (*docked energy*) así como para el conformero de mínima energía perteneciente al *cluster* más poblado.
- GOLD. Con los ocho protocolos descritos en el apartado 5.1.5. En el caso de aplicar la función de *scoring* GOLDScore para guiar el *docking*, no es ésta la función utilizada para calcular la energía de unión, sino el resultado de aplicar la ecuación [1.47], que es para la que se ha obtenido una correlación entre afinidad experimental y valor calculado.<sup>80</sup>


**Figura 7.1.** Scaffold de las estructuras de la Tabla 7.1.

**Tabla 7.1.** Estructuras de los 27 inhibidores considerados en el estudio de afinidad. R<sup>1</sup>, R<sup>2</sup> y R<sup>3</sup> se refieren a los tres puntos de sustitución del scaffold mostrado en la Figura 7.1. nt: no testado.

Nº	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	IC <sub>50</sub> (µM)		
				PDGFR	FGFR	EGFR
1	NH <sub>2</sub>	2,6-diClPhe	Me	4.9	1.3	5.6
2	NHAc	2,6-diClPhe	Me	5.2	5.8	nt
3	NHMe	2,6-diClPhe	Me	6.2	2.5	7.5
4	NHEt	2,6-diClPhe	Me	4.0	0.96	4.5
5	NH- <i>i</i> -Pr	2,6-diClPhe	Me	3.2	0.70	>5
6	NH- <i>n</i> -Bu	2,6-diClPhe	Me	10.3	4.48	4.4
7	NH- <i>c</i> -Hex	2,6-diClPhe	Me	12.5	6.8	>50
8	NH(CH <sub>2</sub> ) <sub>2</sub> OH	2,6-diClPhe	Me	6.5	1.66	>5
9	NHCH <sub>2</sub> Ph	2,6-diClPhe	Me	23.8	14.0	>50
10	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-diClPhe	Me	8.9	13.9	1.4
11	NH(CH <sub>2</sub> ) <sub>3</sub> Nmorfo <sup>a</sup>	2,6-diClPhe	Me	6.8	1.41	>5
12	NH(CH <sub>2</sub> ) <sub>3</sub> Npip(Me) <sup>b</sup>	2,6-diClPhe	Me	2.8	0.38	8.7
13	NH(CH <sub>2</sub> ) <sub>4</sub> Npip(Me) <sup>b</sup>	2,6-diClPhe	Me	0.89	0.32	0.21
14	NH(CH <sub>2</sub> ) <sub>5</sub> Npip(Me) <sup>b</sup>	2,6-diClPhe	Me	0.71	0.24	4.5
15	NHCH <sub>2</sub> -2-piridil	2,6-diClPhe	Me	4.8	3.21	>5
16	NHCH <sub>2</sub> -3-piridil	2,6-diClPhe	Me	6.6	3.15	>5
17	NH(CH <sub>2</sub> ) <sub>2</sub> -2-piridil	2,6-diClPhe	Me	9.3	6.53	>5
31	NHPh(3-OH)	2,6-diClPhe	Me	0.41	0.18	0.15
32	NHPh(4-OH)	2,6-diClPhe	Me	0.28	0.15	nt
33	NHPh(3-CH <sub>2</sub> OH)	2,6-diClPhe	Me	0.21	0.08	0.09
39	NHPh(4-O(CH <sub>2</sub> ) <sub>2</sub> NH <sub>2</sub> )	2,6-diClPhe	Me	0.074	0.055	nt
41	NHPh(4-O(CH <sub>2</sub> ) <sub>2</sub> NMe <sub>2</sub> )	2,6-diClPhe	Me	0.096	0.034	nt
42	NHPh(4-O(CH <sub>2</sub> ) <sub>3</sub> NMe <sub>2</sub> )	2,6-diClPhe	Me	0.12	0.033	nt
43	NHPh(3-O(CH <sub>2</sub> ) <sub>2</sub> NEt <sub>2</sub> )	2,6-diClPhe	Me	0.103	0.032	nt
44	NHPh(4-O(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub> )	2,6-diClPhe	Me	0.079	0.043	0.044
45	NHPh(4-O(CH <sub>2</sub> ) <sub>2</sub> NEt <sub>2</sub> )	2,6-diClPhe	Et	0.15	0.030	nt
51	NHPh(3-COOH)	2,6-diClPhe	Me	0.13	0.11	nt

Por otra parte se aplica el método MM-PBSA/MM-GBSA frente a la estructura de 2fgi. Como se describe en el apartado 1.4.2, este método calcula la energía libre de unión para una serie de *snapshots* obtenidos por dinámica molecular. En este trabajo se realiza para cada ligando una única dinámica molecular de todo el complejo, analizándose dos condiciones diferentes de solvatación (apartado 1.3.6.2), cada uno de los cuales se realiza con diferentes versiones del paquete AMBER, en función de la disponibilidad del mismo:

- Solvente explícito (agua) en condiciones periódicas de contorno con AMBER7. En adelante referida como DIN-SOLEX.
- Solvente explícito en forma de capa de aguas (*cap*) con una corrección para el campo de reacción de las aguas, sin aplicar condiciones periódicas de contorno. Realizada con AMBER8. En adelante referida como DIN-CAP.

El primer esquema requiere más recursos computacionales, aunque conlleva un tratamiento más preciso del solvente y el uso del método PME para el tratamiento electrostático (apartado 1.3.5.1). El último esquema es ampliamente aplicado en este tipo de estudios.<sup>121,125,220</sup> En función de uno u otro esquema, varía el esquema de equilibrado de la dinámica molecular.

El procedimiento general de cálculo MM-PBSA/MM-GBSA conlleva los siguientes pasos:

- Parametrización del ligando. Con el módulo ANTECHAMBER y el *force field* GAFF. Como método de asignación de cargas parciales se escoge AM1-BCC (apartado 1.1.1). En AMBER7 las cargas AM1-BCC se determinan usando el módulo ANTECHAMBER en conexión con MOPAC. AMBER8 dispone de un programa de cálculo semiempírico (DIVCON) que no necesita de MOPAC para el cálculo de las cargas AM1-BCC. Aunque tradicionalmente se exploran las cargas RESP, el método AM1-BCC se comporta adecuadamente, ahorrando tiempo de cálculo.<sup>459</sup>
- Construcción del complejo proteína-ligando. Cada uno de los 27 ligandos se introduce en la conformación resultante del alineamiento flexible de dicho ligando frente a la estructura de PD173074 (17, Figura 5.1) co-cristalizada con FGFR en 2fgi. Para ello, se programa una *script* en SVL que dentro del programa MOE genere el alineamiento de la base de datos inicial de ligandos y seleccione el confórmero de mínima función objetivo (S) de todos los generados para cada uno de los ligandos. Se establece un número máximo de 500 iteraciones y 50 fallos en el alineamiento flexible. Se construye el fichero *pdb* del complejo proteína-ligando. En el caso de DIN-SOLEX, se incluyen también las aguas cristalográficas de 2fgi.

Respecto al *force field* a utilizar para el tratamiento de la proteína, también existen diferencias entre AMBER7 y AMBER8. El primero utiliza el *force field* AMBER Parm99 (ff99), mientras que AMBER8 amplía al anterior con la parametrización de Duan<sup>460</sup> (ff03), con cambios básicos en las torsiones phi y psi.

- Minimización del complejo. En las dos simulaciones se minimiza el complejo con el módulo SANDER durante 10000 iteraciones y con un *cutoff* de 12 Å.
- Solvatación del sistema, equilibrado y etapa de producción de la dinámica.

DIN-SOLEX. Todas estas simulaciones se ejecutan mediante el módulo PMEMD (*Particle Mesh Ewald Molecular Dynamics*), restringido a este tipo de simulación en condiciones periódicas, y no con el módulo SANDER, ya que su velocidad de cálculo está optimizada.

- ◆ Solvatación del sistema con una caja de aguas (WATBOX216) de 12 Å entre la pared de la caja y el átomo más próximo de soluto (complejo).
- ◆ Minimización del sistema durante 1000 iteraciones con un *cutoff* de 12 Å.
- ◆ Equilibrado del sistema. En todas las simulaciones del trabajo se aplica un *time step* de 2 fs aplicando las *restraints* del algoritmo SHAKE. El *cutoff* es de 12 Å:
  - Dinámica molecular de 50 ps con el *ensemble* NVT para equilibrar la temperatura hasta 300 K, con el algoritmo *weak-coupling* (apartado 1.3.3.1) y una constante de acoplamiento  $\tau$  de  $0.2 \text{ ps}^{-1}$ . Se aplican *restraints* a los átomos de la cadena principal de la proteína (C, CA, N, O) y el ligando con una constante de fuerza de 10 kcal/mol (apartado 1.3.7).
  - Dinámica molecular de 50 ps con el *ensemble* NPT para equilibrar la densidad a presión constante y con temperatura de 300 K. Se aplica una constante de acoplamiento de  $0.2 \text{ ps}^{-1}$  con un escalado isotrópico de la presión (apartado 1.3.3.2). Se aplican *restraints* a los átomos de la cadena principal de la proteína y el ligando con una constante de fuerza de 10 kcal/mol (apartado 1.3.7).

- Tres etapas sucesivas de equilibrado con el *ensemble* NVE a 300 K con las mismas *restraints* anteriores con constantes de fuerza de 10 (durante 20 ps), 5 (20 ps) y sin *restraints* (10 ps).

- ◆ Etapa de producción. Durante 500 ps bajo el *ensemble* NVE a 300 K. Sin *restraints*. En principio, se desactiva el algoritmo de acoplamiento de la temperatura. Sin embargo, en algún caso se encuentran inestabilizaciones graves durante la dinámica, por lo que para mantener la temperatura constante a la par que se simula un *ensemble* NVE (sistema más real), se activa dicho algoritmo, aunque aplicándose una constante de acoplamiento más elevada (2.0-5.0 ps<sup>-1</sup>).

DIN-CAP. En este caso, las simulaciones se realizan con el módulo SANDER.

- ◆ Se solvata el sistema con una capa de aguas (*cap*) equilibradas modelo TIP3PBOX centrada entorno al ligando, con un radio de 40 Å (engloba a todo el complejo) y una proximidad máxima de 1 Å entre el soluto y el solvente. Durante las simulaciones esta *cap* se restringe con un potencial semi-harmónico de constante de fuerza de 1kcal/mol en el contorno exterior. En todas ellas, se aplican las condiciones del modelo alternativo de solvente continuo (*Poisson-Boltzmann dynamics*)
- ◆ Minimización de las aguas del sistema durante 2000 iteraciones con un *cutoff* de 30 Å. El resto del sistema se mantiene fijo mediante la aplicación de un potencial harmónico con una constante de fuerza de 500 kcal/mol.
- ◆ Equilibrado del sistema. Con un *cutoff* de 30 Å
  - Dinámica Molecular de 50 ps para calentar el sistema hasta 300 K mediante una rampa con el algoritmo *weak-coupling* y una constante de acoplamiento  $\tau$  de 1 ps<sup>-1</sup>. Se aplican *restraints* a los átomos de la cadena principal de la proteína (C, CA, N, O) y el ligando con una constante de fuerza de 5 kcal/mol.
  - Dinámica Molecular de 20 ps a 300 K con el algoritmo *weak-coupling* y una constante de acoplamiento  $\tau$  de 1 ps<sup>-1</sup>. Sin *restraints*.
- ◆ Etapa de producción. Durante 200 ps a 300 K. Sin *restraints*.
- Cálculo de energías libres de unión mediante MM-PBSA y MM-GBSA. A partir de las trayectorias de la etapa de producción se extraen *snapshots* cada 10 ps para introducir en el cálculo de  $V_{MM}$  y  $G_{solv}$  (ecuación [1.57]), resultando en 50 *snapshots* para DIN-SOLEX y 20 *snapshots* para DIN-CAP. El solvente explícito se elimina de estos *snapshots* para realizar el cálculo, ya que en este punto se emplea solvente implícito (apartado 1.3.6.3).

El término de solvatación se calcula con el método generalizado de *Born* (MM-GBSA) (apartado 1.3.6.3.2) para DIN-SOLEX y DIN-CAP. Además, para DIN-CAP se calcula dicho término con la versión implementada en AMBER8 de resolución de la ecuación de Poisson-Boltzmann (apartado 1.3.6.3.1). En cada caso, se ajustan los parámetros de  $\gamma$  y  $\beta$  para el cálculo del término  $\Delta G_{SA}$  (ecuación [1.63]) a los correspondientes mostrados en la Tabla 1.1. La constante dieléctrica dentro y fuera del soluto es de 1 y 80, respectivamente.

El término entrópico se calcula mediante análisis de modos normales (módulo NMODE de AMBER) únicamente para DIN-CAP. Debido a los requerimientos computacionales de memoria para el cálculo de la entropía,  $S_{proteína}$  o  $S_{complejo}$ , no es posible realizar dicho cálculo para la totalidad del sistema. Por ello, se extraen los residuos comprendidos en una esfera de 12 Å centrados sobre el ligando para cada uno de los *snapshots*.

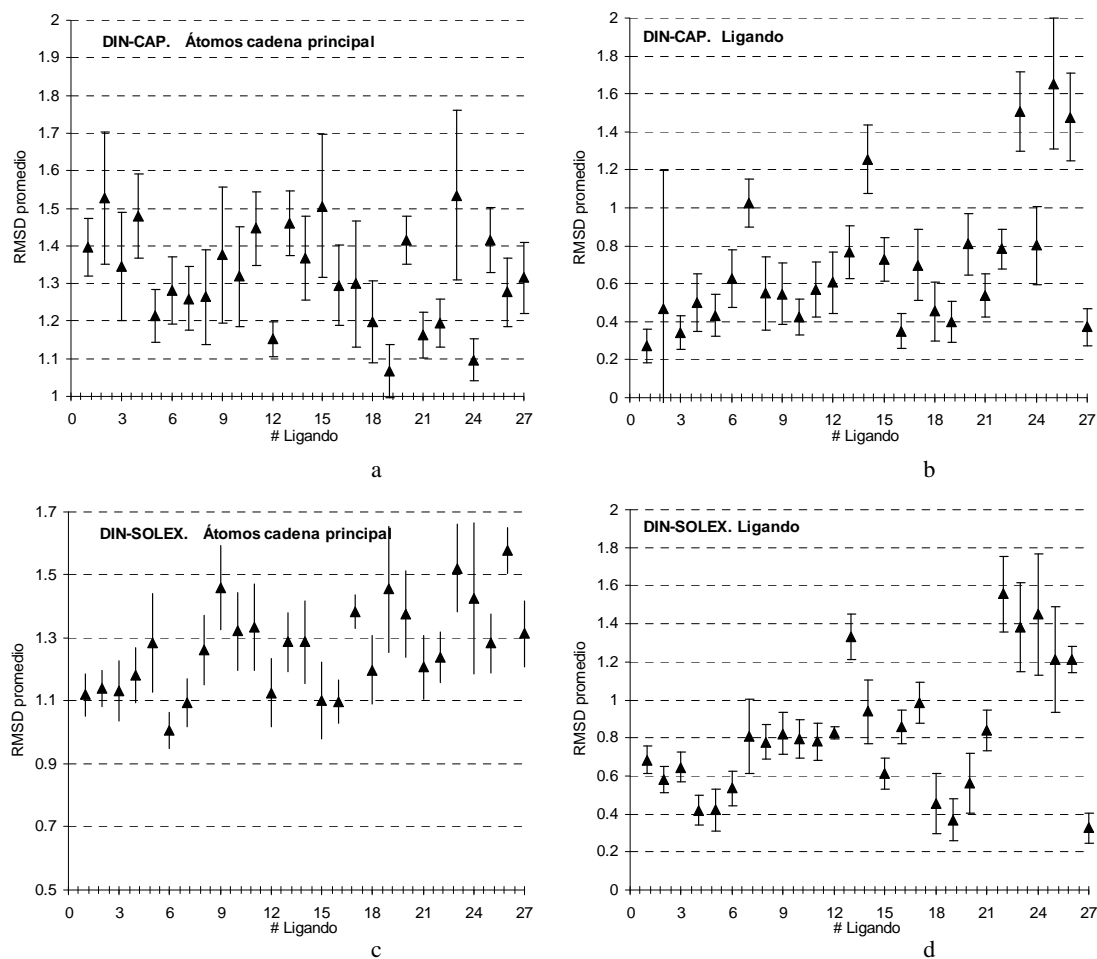
Además, dado el elevado tiempo computacional, únicamente se usan *snapshots* extraídos cada 40 ps para DIN-CAP, esto es un total de 5 *snapshots*. Para cada uno de estos *snapshots*, se calcula la entropía para la estructura minimizada en vacío con una constante dieléctrica dependiente de la distancia ( $\epsilon=4r$ ).

Para realizar todo este proceso (acondicionamiento del sistema, simulaciones, cálculo MM-PBSA) se prepara una *script* en perl que toma como entrada el *pdb* del ligando, amén de disponer del *pdb* de la proteína y aguas cristalográficas así como de los ficheros de entrada con las especificaciones de las simulaciones. Esta *script* modifica los nombres de los átomos del ligando de manera que sean diferentes, requisito imprescindible para ANTECHAMBER.

El procesado de los *snapshots* para extraer los residuos de interés en el cálculo de NMODE se realiza mediante *scripts* implementadas en perl y en lenguaje SVL dentro de MOE.

### 7.1.2. Resultados del cálculo de energías libres de unión

Para las dos simulaciones, se calcula mediante el módulo CARNAL la RMSD de los átomos de la proteína (de todos los átomos y de los de la cadena principal) y del ligando a lo largo de la etapa de producción, referidos a la estructura del complejo minimizada. En la Figura 7.2 se muestran las RMSD promedio de los átomos de la cadena principal (C, CA, N, O) de la proteína (y su desviación estándar) y la RMSD promedio (y desviación estándar) de los inhibidores calculadas en las 27 simulaciones con DIN-CAP y 27 simulaciones con DIN-SOLEX. Las pequeñas fluctuaciones, así como la estabilización energética del sistema (no mostrado), indican que el sistema se encuentra equilibrado y estable.



**Figura 7.2.** RMSD promedio y desviación estándar de (a) los átomos de la cadena principal con DIN-CAP, (b) los ligandos con DIN-CAP, (c) los átomos de la cadena principal con DIN-SOLEX y (d) los ligandos con DIN-SOLEX, para las 27 simulaciones.

Se calcula el promedio de la ocupación del puente de hidrógeno entre Ala<sup>564</sup> y el correspondiente átomo de nitrógeno del inhibidor a lo largo de la fase de producción de cada una de las dinámicas. Para DIN-CAP, en 23 de las 27 simulaciones, el promedio supera el 90% a lo largo de la dinámica, correspondiendo la mínima ocupación al inhibidor 17 (Tabla 7.1), con un 41% de ocupación. Para DIN-SOLEX, de las 27 simulaciones, en 25 el promedio de ocupación supera el 90% a lo largo de la dinámica.

En las Tablas 7.2 y 7.3 se muestran los componentes de la energía libre para DIN-CAP y DIN-SOLEX, respectivamente, calculados con MM-GBSA y MM-PBSA.

**Tabla 7.2.** Componentes de la Energía Libre para **DIN-CAP** en kcal/mol.

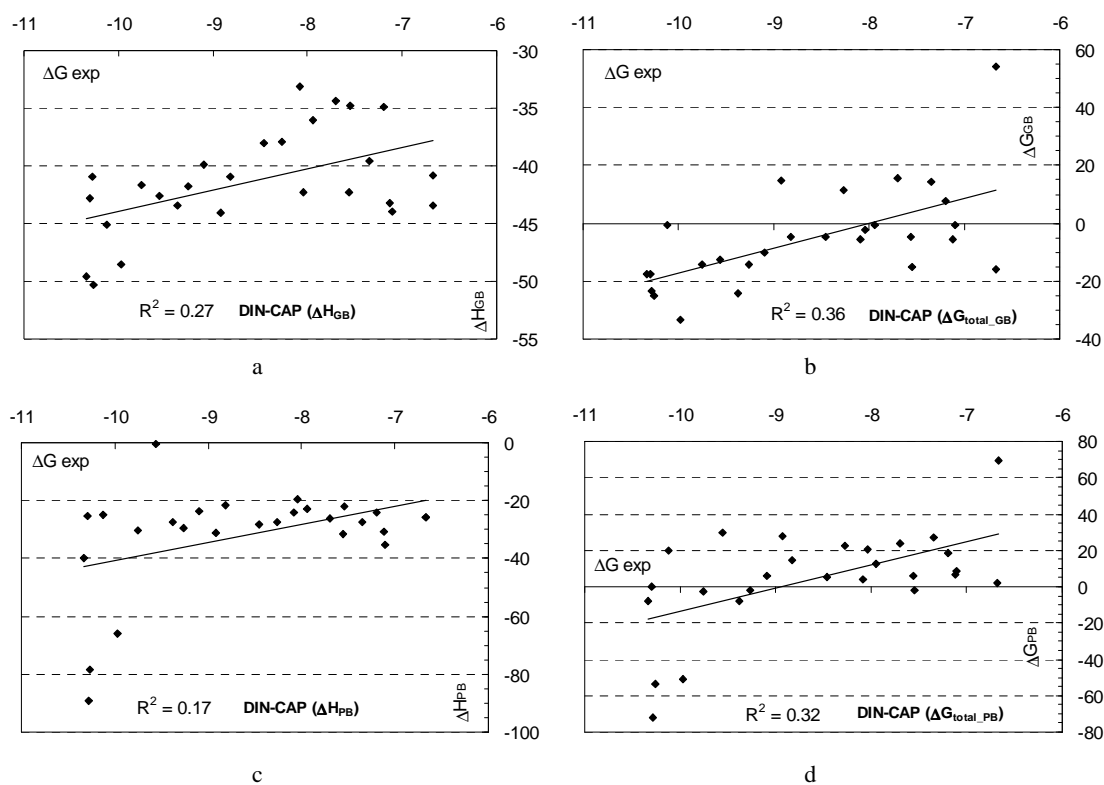
N	$\Delta H_{gas}$	MM-GBSA			MM-PBSA			-TAS	$\Delta G$		$\Delta G_{exp}$
		$\Delta G_{SA}$	$\Delta G_{GB}$	$\Delta H_{GB}$	$\Delta G_{SA}$	$\Delta G_{PB}$	$\Delta H_{PB}$		total_GB	total_PB	
1	-55.81	-4.89	27.58	-33.12	-3.40	35.25	-23.96	27.72	-5.40	3.76	-8.09
2	-60.47	-5.60	31.15	-34.92	-3.89	40.44	-23.92	42.50	7.58	18.58	-7.19
5	-58.07	-5.13	28.83	-34.37	-3.56	35.61	-26.02	49.84	15.47	23.82	-7.70
6	-58.28	-5.39	25.74	-37.93	-3.74	34.70	-27.32	49.59	11.66	22.27	-8.27
7	-58.11	-5.56	25.66	-38.01	-3.86	33.63	-28.34	33.44	-4.57	5.10	-8.46
8	-58.92	-5.83	25.18	-39.57	-4.05	35.72	-27.25	54.13	14.56	26.88	-7.35
9	-68.87	-6.04	30.91	-44.00	-4.20	37.95	-35.12	43.41	-0.59	8.29	-7.10
10	-62.52	-5.51	31.99	-36.04	-3.82	43.33	-23.01	35.64	-0.40	12.63	-7.94
11	-67.71	-5.89	32.78	-40.82	-4.10	46.08	-25.73	94.96	54.17	69.26	-6.67
13	-68.90	-6.65	32.13	-43.42	-4.62	47.91	-25.61	27.52	-15.90	1.91	-6.67
14	-66.37	-6.71	30.74	-42.34	-4.66	51.66	-19.37	40.04	-2.30	20.67	-8.04
15	-67.05	-6.40	32.50	-40.95	-4.44	49.75	-21.74	36.09	-4.86	14.35	-8.82
16	-70.66	-6.49	33.09	-44.06	-4.51	44.02	-31.15	58.92	14.86	27.77	-8.92
17	-71.35	-6.87	38.35	-39.87	-4.77	52.50	-23.62	29.75	-10.12	6.13	-9.10
18	-58.89	-5.84	29.95	-34.78	-4.06	41.13	-21.82	19.84	-14.94	-1.98	-7.55
19	-71.52	-5.85	35.04	-42.33	-4.06	44.10	-31.48	37.46	-4.87	5.98	-7.56
20	-67.40	-6.07	30.28	-43.19	-4.22	40.93	-30.69	37.58	-5.61	6.89	-7.12
31	-70.36	-5.75	34.32	-41.79	-4.00	44.97	-29.39	27.52	-14.27	-1.87	-9.27
32	-68.96	-6.04	31.55	-43.45	-4.20	45.97	-27.19	19.17	-24.28	-8.02	-9.38
33	-71.35	-6.10	35.83	-41.62	-4.24	45.27	-30.32	27.39	-14.23	-2.93	-9.75
39	-82.48	-6.46	40.40	-48.54	-4.48	20.84	-66.12	15.04	-33.50	-51.08	-9.97
41	-70.83	-6.60	27.11	-50.32	-4.59	-3.19	-78.61	25.10	-25.22	-53.51	-10.26
42	-42.52	-6.10	7.70	-40.92	-4.24	-42.35	-89.11	17.33	-23.59	-71.78	-10.28
43	-65.42	-7.20	29.77	-42.85	-5.00	45.07	-25.35	25.14	-17.71	-0.21	-10.30
44	-74.05	-6.65	35.55	-45.15	-4.62	53.63	-25.04	44.58	-0.57	19.54	-10.12
45	-82.22	-6.77	39.36	-49.63	-4.70	47.29	-39.63	31.84	-17.79	-7.79	-10.34
51	-155.9	-6.29	119.6	-42.62	-4.37	159.8	-0.43	30.07	-12.55	29.64	-9.56

En las Figuras 7.3 (DIN-CAP) y 7.4 (DIN-SOLEX) se representa  $\Delta H$  (sin considerar entropía) y  $\Delta G$  calculados usando MM-GBSA/MM-PBSA (eje y) frente a la energía experimental (eje x). Los resultados con DIN-CAP son bastante pobres, mientras que con DIN-SOLEX se obtiene una mejor correlación ( $R^2=0.51$ ). Esta mejora puede ser debida al tratamiento más riguroso del sistema así como al mayor tiempo de simulación de DIN-SOLEX frente a DIN-CAP (500 ps vs 200 ps).

Se observa también cómo la inclusión del término entrópico conduce a energías de unión positivas (Figura 7.3.a vs 7.3.b y 7.3.c vs 7.3.d). Aunque evidentemente no tiene significación física, resulta más relevante el hecho de que son capaces de aumentar la predicción del orden relativo de las moléculas. Por otra parte, las energías de solvatación  $\Delta G_{GB}$  y  $\Delta G_{PB}$  aunque deberían ser comparables por ambos métodos, no lo son. En este sentido, el método generalizado de Born rinde mejores predicciones que Poisson-Boltzmann.

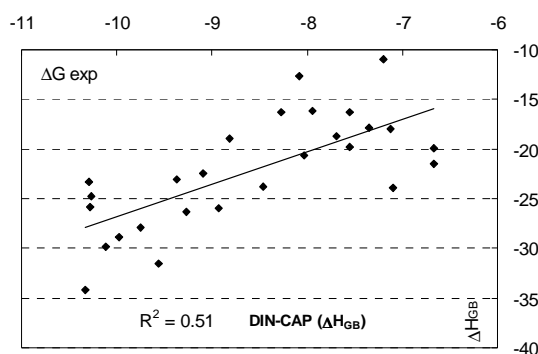
**Tabla 7.3.** Componentes de la Energía Libre para **DIN-SOLEX** en kcal/mol.

N	$\Delta H_{gas}$	MM-GBSA			$\Delta G_{exp}$
		$\Delta G_{SA}$	$\Delta G_{GB}$	$\Delta H_{GB}$	
1	-50.84	-3.32	41.54	-12.62	-8.09
2	-56.91	-4.10	49.99	-11.02	-7.19
5	-53.06	-3.35	37.71	-18.70	-7.70
6	-57.10	-3.86	44.62	-16.35	-8.27
7	-61.14	-3.62	41.02	-23.74	-8.46
8	-56.72	-3.48	42.32	-17.88	-7.35
9	-63.79	-3.93	43.81	-23.91	-7.10
10	-50.72	-4.10	38.62	-16.20	-7.94
11	-61.20	-4.25	44.00	-21.45	-6.67
13	-59.26	-4.11	43.50	-19.87	-6.67
14	-65.93	-4.83	50.14	-20.62	-8.04
15	-60.34	-4.01	45.35	-19.00	-8.82
16	-69.63	-4.34	48.00	-25.97	-8.92
17	-63.15	-4.38	45.09	-22.44	-9.10
18	-54.48	-3.71	41.84	-16.35	-7.55
19	-65.18	-4.29	49.67	-19.80	-7.56
20	-67.99	-4.24	54.21	-18.02	-7.12
31	-75.25	-3.99	52.94	-26.30	-9.27
32	-65.62	-3.97	46.48	-23.11	-9.38
33	-72.55	-4.06	48.68	-27.93	-9.75
39	-76.56	-4.15	51.84	-28.87	-9.97
41	-69.06	-4.29	48.64	-24.71	-10.26
42	-67.86	-4.37	46.40	-25.83	-10.28
43	-63.20	-5.23	45.07	-23.36	-10.30
44	-74.43	-4.86	49.50	-29.79	-10.12
45	-84.13	-5.03	54.94	-34.22	-10.34
51	-141.6	-3.98	114.0	-31.58	-9.56



**Figura 7.3.** Energías libres de unión ( $\Delta G$ ) y entalpías libres de unión ( $\Delta H$ ) para cada inhibidor de FGFR calculadas con MM-PBSA y MM-GBSA a partir de los *snapshots* extraídos de **DIN-CAP** versus la energía experimental. (a)  $\Delta H$  calculada con MM-GBSA. (b)  $\Delta G$  calculada con MM-GBSA. (c)  $\Delta H$  calculada con MM-PBSA. (d)  $\Delta G$  calculada con MM-PBSA.





**Figura 7.4.** Entalpía libre de unión ( $\Delta H$ ) para cada inhibidor de FGFR calculada con MM-GBSA a partir de los *snapshots* extraídos de DIN-SOLEX *versus* la energía experimental.

Finalmente se muestran en la Tabla 7.4 los coeficientes de correlación obtenidos con las diferentes condiciones de MM-PBSA/MM-GBSA y las funciones de *scoring* del *docking*. Los valores obtenidos para la simulación DIN-CAP con MM-PBSA son del orden de aquellos calculados usando GOLDScore y CHEMSCore, siendo estos últimos métodos mucho más eficientes computacionalmente. DIN-SOLEX alcanza un coeficiente de  $R_s$  de 0.69 (PI de 0.72), el máximo de los obtenidos. Para DIN-CAP, la inclusión del término entrópico, aunque conduce a valores positivos de  $\Delta G$ , conduce a un ranking de las moléculas más similar al ranking experimental. Quedaría por ver si la inclusión del término entrópico en DIN-SOLEX permitiría incrementar el coeficiente de correlación, cálculo no realizado por una limitación de tiempo.

Estos coeficientes de correlación obtenidos con el método MM-PBSA se sitúan en un intermedio de los reportados en otros estudios con este método. En el estudio de Pearlman<sup>121</sup> de 16 inhibidores tipo 1,7,8-triaza-naftalen-6-ona de la quinasa p38 se obtiene un PI máximo del orden de 0.51 con un tiempo de simulación de 5 ns calculados sobre dinámicas separadas para complejo, proteína y ligando, que disminuye considerablemente al reducirse el tiempo de simulación a 1 ns o 250 ps con una única dinámica (PI de 0.15).<sup>125</sup> Al aplicarse este método a otro set de 12 inhibidores tipo pirido[2,3-*d*]pirimidinas y otros *scaffolds* de la quinasa p38 se obtiene un PI de 0.37 para una simulación de 250 ps con una única dinámica.<sup>125</sup> Una de las mejores predicciones se ha obtenido frente a 8 ligandos de avidina, con valores de PI de 0.86<sup>125</sup> y 0.99<sup>122</sup>, en función de las condiciones de cálculo aplicadas.

Tal y como sugieren estos estudios, es esperable que las predicciones mejorasen al incrementarse el tiempo de simulación de 200/500 ps a varios nanosegundos, aunque cabría ver si la mejora justifica el tiempo computacional añadido. Además, la bondad de las predicciones depende del margen de actividades estudiado. Es decir, como puede observarse en las Figuras 7.3 y 7.4, si bien el método discrimina entre ligandos fuertes y débiles, no discierne tan bien entre aquellos con pequeñas diferencias de  $\Delta G$  (1-2 kcal/mol).

**Tabla 7.4.** Coeficientes de correlación ( $R_s$ , PI,  $R^2$ ) obtenidos para los diferentes métodos de *scoring* con las funciones de *docking* y el método MM-PBSA/MM-GBSA.

Método de <i>scoring</i>	$R_s$	PI	$R^2$
AUTODOCK Gasteiger-Marsili 101×101×101 Confórmero de mínima energía	-0.02	0.01	0.01
AUTODOCK Gasteiger-Marsili 101×101×101 Confórmero del <i>cluster</i> más poblado	0.27	0.29	0.07
AUTODOCK Gasteiger-Marsili 61×61×61 Confórmero de mínima energía	0.37	0.40	0.16
AUTODOCK Gasteiger-Marsili 61×61×61 Confórmero del <i>cluster</i> más poblado	0.16	0.19	0.04
AUTODOCK AM1 101×101×101 Confórmero de mínima energía	0.46	0.48	0.21
AUTODOCK AM1 101×101×101 Confórmero del <i>cluster</i> más poblado	0.17	0.19	0.01
AUTODOCK AM1 61×61×61 Confórmero de mínima energía	0.43	0.44	0.18
AUTODOCK AM1 61×61×61 Confórmero del <i>cluster</i> más poblado	0.48	0.49	0.20
GOLDScore	0.24	0.46	0.19
GOLDScore-RESCORE	0.26	0.51	0.10
GOLDScore-RESTRAINT	0.51	0.53	0.21
GOLDScore-RESTRAINT-RESCORE	0.45	0.46	0.23
CHEMScore	0.39	0.44	0.13
CHEMScore-RESCORE	0.54	0.57	0.20
CHEMScore-RESTRAINT	0.20	0.23	0.05
CHEMScore-RESTRAINT-RESCORE	0.24	0.26	0.07
DIN-CAP ( $\Delta H_{GB}$ )	0.43	0.46	0.27
DIN-CAP ( $\Delta H_{PB}$ )	0.25	0.26	0.17
DIN-CAP ( $\Delta G_{total\_GB}$ )	0.58	0.59	0.36
DIN-CAP ( $\Delta G_{total\_PB}$ )	0.49	0.50	0.32
DIN-SOLEX ( $\Delta H_{GB}$ )	0.69	0.72	0.51

## 7.2. Aplicación de MM-PBSA en cribado virtual

### 7.2.1. Procedimiento para el cribado virtual

En este apartado se analiza la aplicación del método MM-GBSA en cribado virtual, aplicación recientemente introducida por Kuhn<sup>125</sup>. Dado su elevado coste computacional, no es factible realizar una dinámica molecular para todos los ligandos contenidos en una base de datos del tamaño de las usadas en este estudio (~5000 compuestos). El total de las simulaciones para cada ligando (sin contar el cálculo de MM-PBSA) es de unas 95-100 horas en un Pentium IV a 2.4GHz. Así, tal y como Kuhn propone, es más adecuada su aplicación como filtro secundario tras el *docking*, de manera que se analice una única conformación (*snapshot*) por compuesto, actuando el método MM-GBSA únicamente como función de *scoring* sin muestrear el espacio conformacional. Como se indica en la introducción, esta novedosa aplicación no deja de ser controvertida, ya que en principio este método operaría mejor con un amplio rango de conformaciones (*snapshots*) a analizar extraídas de una o varias dinámicas moleculares.

Sin embargo, Kuhn y colaboradores encontraron que en ciertos casos el protocolo utilizando una única conformación minimizada rinde factores de enriquecimiento superiores a los encontrados en un protocolo con una dinámica molecular de 200 ps (50 ps de etapa de producción).<sup>125</sup>

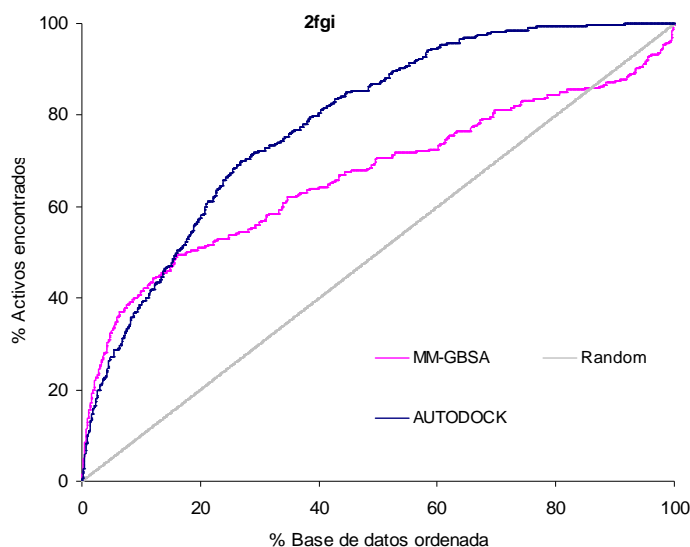
En este caso, se escoge la base de datos Base\_Maybridge, formada por 288 inhibidores de FGFR y 4342 inactivos, tomando únicamente la conformación para cada ligando de menor *docked energy* resultante del *docking* con AUTODOCK frente a 2fgi (FGFR) (apartado 5.2.4).

El procedimiento es análogo al descrito en el apartado 7.1.1, aunque en lugar de realizar una dinámica molecular, una vez construido y solvatado el complejo mediante una capa de aguas, se realiza una minimización de 1000 iteraciones con un *cutoff* de 30 Å, sin ningún tipo de restricción. Se utiliza AMBER8, que dispone de un *force field* GAFF más completo, calculándose las cargas AM1-BCC con el programa DIVCON. Una vez obtenidas las coordenadas del complejo éstas se transforman a formato de trayectoria molecular mediante el módulo PTRAJ. El término de energía de solvatación se calcula únicamente con el método generalizado de Born (MM-GBSA), más eficiente computacionalmente.

Siguiendo este protocolo, aproximadamente, el tiempo (*wallclock*) para procesar un ligando es de unos 10-15 minutos en un procesador Pentium IV a 2.4 GHz bajo Linux.

## 7.2.2. Resultados del cribado virtual con MM-GBSA

En la Figura 7.5 se muestra la curva de enriquecimiento obtenida con el cribado virtual relativa a una selección aleatoria. Para 502 de las 4630 moléculas (10.8%) no se puede calcular MM-PBSA debido a la falta de parametrización de los radios para el cálculo de la superficie accesible de los átomos F, Br, I. De estas 502 moléculas, 33 son compuestos activos. Así, los factores de enriquecimiento se calculan sobre la base de 255 activos en un total de 4128 compuestos. Por comparación, se incluye también la curva de enriquecimiento obtenida para Base\_Maybridge con AUTODOCK (equivale a la mostrada en la Figura 5.17.d), que contiene 288 activos en un total de 4630 compuestos.



**Figura 7.5.** Curva de enriquecimiento obtenida con el protocolo MM-GBSA frente a FGFR para la estructura 2fgi en el cribado de Base\_Maybridge. La bisectriz (línea gris) representa los valores obtenidos si los activos se seleccionasen aleatoriamente.

En la Tabla 7.5 se muestran los factores de enriquecimiento (*ef*), en el primer 2%, 5% y 10% de base de datos Base\_Maybridge cribada con MM-GBSA y AUTODOCK. Dadas las diferencias en el número de compuestos de cada base de datos, se muestra junto a *ef* el número de compuestos cribados totales (*N*) que supone cada porcentaje.

**Tabla 7.5.** Factores de enriquecimiento (*ef*) obtenidos en el primer 2%, 5% y 10% de base de datos cribada frente a FGFR (2fgi). *N* corresponde al número de compuestos totales cribados a cada porcentaje para cada uno de los dos métodos.

	2%		5%		10%	
	<i>N</i>	<i>ef</i>	<i>N</i>	<i>ef</i>	<i>N</i>	<i>ef</i>
MM-GBSA	83	9.9	207	6.5	413	4.1
AUTODOCK	93	8.0	232	5.4	463	3.9

Como se observa en la Figura 7.5 y la Tabla 7.5, MM-GBSA identifica un mayor número de activos que AUTODOCK en los primeros porcentajes de la base de datos cribada, aunque su comportamiento empeora para ciertos compuestos, apareciendo éstos a porcentajes más elevados. Se lista en la Tabla 7.6 el porcentaje de base de datos cribados a los que se recupera al menos un *lead* representativo de cada uno de los seis *scaffolds* contenidos en el *pool* de activos. Básicamente, la principal diferencia se debe a la identificación de un representante de los fenilbenzimidazoles, que AUTODOCK posiciona en un rango elevado, aunque un análisis visual de este primer compuesto recuperado demuestra que no se encuentra bien orientado en el sitio de unión.

**Tabla 7.6.** Porcentaje de base de datos cribada frente a FGFR para la que al menos un representante activo de cada uno de los seis *scaffolds* de Base\_Maybridge es recuperado.

	2fgi					
	PD	Naph	Urea	Indol	Phen	Anil
MM-GBSA	0.1	0.02	0.05	0.8	0.4	<b>4.4</b>
AUTODOCK	0.02	0.2	0.1	0.9	<b>30.8</b>	2.3

*PD*: pirido[2,3-*d*]pirimidinas, *Naph*: naftiridin-2(1*H*)-onas, *Urea*: 7-alquilurea pirido[2,3-*d*]pirimidinas, *Indol*: indolin-2-onas, *Fenil*: 1-fenilbenzimidazoles y *Anilino*: 4-anilinoquinazolininas. Los números en negrita corresponden al peor caso, es decir, el porcentaje de base de datos ordenada necesario para recuperar al menos un representante de cada uno de los *cores*.