



**Universitat Ramon Llull**

## **TESIS DOCTORAL**

Título	<b>Herramientas de cribado virtual aplicadas a inhibidores de tirosina quinasa. Contribución al desarrollo del programa PRALINS para el diseño de quimiotecas combinatorias.</b>
Realizada por	Obdulia Rabal Gracia
en el Centro	Escola Tècnica Superior IQS
y en el Departamento	Química Orgánica y Bioquímica
Dirigida por	Dr. Jordi Teixidó i Closa



## Capítulo 8.

### PRALINS: Program for Rational Analysis of Libraries *in Silico*

Tal y como se expone en la introducción, se ha continuado ampliando el programa PRALINS, programa dirigido al diseño y selección de quimiotecas combinatorias. PRALINS consta tanto de una versión en modo texto (programada en ANSI-C) como de una versión gráfica (librerías *Xt/Motif.Xt Intrinsic*) con un editor y visualizador de moléculas. En la referencia [303] puede encontrarse una descripción detallada de las variadas funcionalidades y las instrucciones para su uso que ofrece la versión original.

En este capítulo se presentan las diversas líneas ampliadas, orientando la exposición a la aplicación de las mismas a diversos casos de estudio. Se detallan las instrucciones necesarias para ejecutar PRALINS en cada caso, programadas únicamente para la versión en modo texto en el que las instrucciones se introducen mediante un fichero ASCII. Las distintas ampliaciones son:

- i) Implementación de algoritmos genéticos (GA) como método de optimización (módulos de similitud y diversidad).
- ii) Métodos de selección de quimiotecas combinatorias focalizadas, introducción de un nuevo método.
- iii) Ampliación de la funcionalidad para evaluar la diversidad de las quimiotecas seleccionadas. Se analizan los diferentes criterios, se propone una nueva alternativa.
- iv) Optimización multiobjetivo con algoritmos genéticos (MOGA).
- v) Otras implementaciones menores, como la ampliación de los métodos de *clustering*.

#### 8.1. Implementación de algoritmos genéticos (GA)

La versión original de PRALINS dispone como métodos heurísticos de optimización para las selecciones basadas en diversidad de: Monte Carlo (MTC), *Simulated Annealing* (SA), búsqueda local (LS), así como la posibilidad de realizar una búsqueda exhaustiva completa (complete). En este punto se describe la implementación de los algoritmos genéticos en PRALINS como método de optimización de selecciones:

- *Full array* (subquimiotecas) focalizadas.
- *Full array* (subquimiotecas) diversas, con métodos basados en distancias, *binning* y *clustering* (jerárquico, Jarvis-Patrick y *K-means*)
- *Cherry picking* o *sparse array* diversas con métodos basados en distancias: MaxMin (ecuación [1.108]), MaxSum (ecuación [1.109]) y MaxMin promediado (ecuación [1.110]).

Los principios básicos de este método heurístico se recogen en el apartado 1.9.2, por lo que la exposición se ciñe a las características particulares en el contexto de selección de quimiotecas.

Cada cromosoma del algoritmo genético representa una posible solución, es decir, un *array* de compuestos para las selecciones *cherry picking* o una lista de reactivos, origen de una quimioteca combinatoria para el diseño *full array*. En este último caso, el cromosoma se subdivide en tantas secciones como *steps* o puntos de variabilidad/sustitución contiene la quimioteca combinatoria.

El primer punto que se plantea es la codificación de la información en el cromosoma. Como se indica en el apartado 1.10.3, varios autores han propuesto diferentes sistemas de codificación:

- Cromosomas cuya longitud es equivalente al número de compuestos a seleccionar y donde cada gen es un entero correspondiente a la identificación de dicho compuesto.<sup>156,157</sup> Presenta la ventaja de que no necesita una función de codificación/descodificación del vector. Su principal desventaja es que obliga a un tamaño de selección fijo, perdiéndose la posibilidad de optimizar esta variable en una optimización multiobjetivo. Implementado exclusivamente y con carácter de prueba para las selecciones *cherry picking* con los criterios MaxMin, MaxSum y MaxMin promediado.
- Cromosomas en los que el número de genes corresponde al número de reactivos/productos disponibles, de manera que cada gen se representa por un bit activado o no en función de si el reactivo/producto está seleccionado o no. En este caso, se requiere una función de codificación/descodificación del cromosoma antes de evaluarse su *fitness* que transforme los valores de enteros a bits y viceversa. El tamaño y configuración (número de reactivos para ser seleccionados para cada *step* de un *full array*) pueden modificarse dinámicamente durante la optimización. En principio, se fija el tamaño/configuración, por lo que el número de genes activados debe mantenerse constante en cada sección del cromosoma en la solución final. Para ello, tras el evento de mutación, si el cromosoma obtenido sobrepasa o no alcanza el número deseado de genes activados, se fuerza a reducir o incrementar aleatoriamente este número hasta alcanzar dicho valor. Implementado para selecciones *cherry picking* y *full array*.

En la Figura 8.1 se muestra un ejemplo de ambas representaciones para la selección hipotética de un *full array* 5×5×5 a partir de una quimioteca combinatoria 10×10×10.

Codificación en enteros:	1 5 3 2 4   2 3 7 9 10   6 3 5 8 7
Codificación en bits:	1111100000   0110001011   0010111100

**Figura 8.1.** Ejemplo de los dos esquemas utilizados para la codificación del cromosoma.

Tras generarse una población inicial de tamaño fijado por el usuario, se calcula el valor de la función objetivo a optimizar. Dicha población puede generarse enteramente de manera aleatoria o permitiendo que una fracción determinada de cromosomas adopte una solución propuesta por el usuario (pero no todas ellas, para evitar una homogenización de las soluciones). Dicha fracción es un parámetro interno del programa, con un valor asignado por defecto de 0.5. El acondicionamiento de la función objetivo según el caso (similitud, diversidad, métrica considerada) se realiza automáticamente, considerando la posibilidad de que la función adopte valores negativos y la necesidad de transformar la función para que se maximizen siempre valores (apartado 1.9.2.3).

En cada iteración se genera una población de descendientes aplicando el operador de selección a un par de padres, que generan un par de hijos, tantas veces como sea necesario para completar el tamaño de población, que se mantiene constante. Nótese que de este modo el tamaño de

población requerido por el usuario debe ser un número par. Como operadores de selección se incorporan en PRALINS el *Roulette Wheel selection* y el *Stochastic Remainder Selection Without Replacement*. En ambos se realiza un escalado lineal de la función de *fitness*, con un valor máximo de la misma de 2.0 (parámetro interno del programa). Una tercera alternativa de método de selección es la de aplicar el operador *Roulette Wheel selection* pero con una ordenación de los valores de la función objetivo proporcional al rango (apartado 1.9.2.3).

Dentro de las posibilidades de operadores de *crossover* (apartado 1.9.2.4) se implementan el *single-point crossover*, el *two-point crossover* y el *uniform crossover*, todos ellos para las tres posibilidades (codificación en enteros, *cherry picking* codificación en bits y *full array* codificación en bits). En el formato *full array* cada evento de *crossover* afecta únicamente a un *step* aleatorio en cada paso. La probabilidad de que suceda un evento de *crossover* viene fijada por la tasa de *crossover*, valor comprendido entre 0 y 1. Se genera un número aleatorio entre 0 y 1 y si éste es inferior a la tasa de *crossover*, el evento tiene lugar.

Tras el *crossover* tiene lugar la mutación para cada uno de los cromosomas hijos generados, también con una probabilidad determinada de forma análoga según una tasa de mutación. Inicialmente, se testan dos versiones del operador mutación en el caso de la codificación por enteros en *cherry picking*. En la primera de ellas, se evalúa la probabilidad de mutación para cada uno de los alelos del cromosoma padre. En la segunda de ellas, únicamente se permite un evento de mutación para cada uno de los cromosomas. El primer esquema conduce a una elevadísima tasa de mutaciones, aún y disminuyendo considerablemente la probabilidad de mutación, conduciendo a soluciones excesivamente alejadas de la inicial y perdiéndose información, por lo que se desestima su aplicación.

Por ello, en la codificación en bits únicamente se aplica la segunda versión. Además, en caso de selecciones *full array* el evento de mutación ocurre sobre el mismo *step* (sección del cromosoma) sobre el que se ha aplicado el evento previo de *crossover*. En caso de que no haya habido *crossover* previo en este mismo evento reproductivo, dicho *step* se selecciona aleatoriamente.

Como modelo de reemplazo (*replacement*) se considera únicamente el modelo generacional (apartado 1.9.2.5, Figura 1.18), ya que otros autores han apuntado que produce mejores soluciones que el modelo *steady-state*, aunque este último converge más rápidamente.<sup>154</sup>

Las pruebas iniciales con un esquema de reemplazo incondicional demostraron que la presión selectiva del método de selección no es suficiente para garantizar una buena convergencia del algoritmo. Por ello se opta por implementar diversos esquemas de reemplazo condicional y elitismo:

- Reemplazo condicional: la nueva generación se compone de los mejores individuos de la generación parental y filial.
- Reemplazo condicional sin repetición: como en el caso anterior, pero impidiendo la repetición de individuos con genotipos idénticos.
- Elitismo. El usuario impone que un determinado número de los mejores cromosomas de la población parental pasen a la siguiente generación. El resto de soluciones se reemplaza por las mejores soluciones de la generación filial.

Estas tres funciones están disponibles para los tres formatos (*cherry picking* con enteros, *cherry picking* con bits y *full array* con bits).

El proceso iterativo termina tras un número definido de generaciones o si el valor óptimo obtenido para la función objetivo no mejora tras un número especificado de pasos. La información del mismo queda recogida en el fichero *pointsbit.txt*, donde se incluyen los valores

de la función objetivo para la suma de individuos de la generación, el promedio de la misma, el mejor y el peor de los individuos en cada una de las iteraciones.

Tanto las selecciones *cherry picking* como *full array* se programan considerando que el usuario pueda fijar en la solución final unos determinados productos o reactivos, que evidentemente afectarán a la medida de diversidad/similitud a optimizar.

Además, las selecciones *full array* se programan con el concepto de grupo, concepto también disponible en otros métodos de optimización. Este concepto puede aplicarse en aquellas quimiotecas en las que dos o más puntos de diversidad o *steps* comparten listas de reactivos candidatos, imponiéndose a la selección *full array* una nueva restricción por la que los reactivos seleccionados para estos *steps* sean iguales o estén incluidos en la lista de los seleccionados para aquel *step* con un mayor número de reactivos requeridos. Si no se especifica esta opción de grupo, cada uno de los *steps* de la quimioteca compone un grupo independiente. La utilidad de esta imposición adicional al criterio *full array* es que permite disminuir aún más el número de reactivos totales necesarios para sintetizar la quimioteca seleccionada, a la par que se maximiza el criterio de diversidad o similitud impuesto. Las quimiotecas peptídicas constituyen un buen ejemplo donde aplicar este concepto.

### 8.1.1. Instrucciones de cálculo en PRALINS con algoritmos genéticos

Esta descripción, así como las que prosiguen, se centra exclusivamente en las funcionalidades introducidas en este trabajo, entendiendo que el resto de opciones para ejecutar el programa en modo texto se encuentran recopiladas en la referencia [303]. Así, debe consultarse esta referencia para encontrar la información necesaria para preparar el fichero de entrada ASCII correspondiente (indicación del fichero *sd-file* de entrada, características de la quimioteca, tamaño y método de selección, métrica, descriptores a considerar, reactivos de partida, reactivos y productos fijados, definición del concepto de grupo...).

PRALINS utiliza un código numérico para cada uno de los métodos de optimización disponibles, incluido en la línea de identificación del método de selección (véase referencia [303]). Para las selecciones con GA codificado en bits este código corresponde a 9, mientras que la codificación en enteros (exclusiva de *cherry picking*) utiliza un 8. En la línea siguiente deben especificarse los siguientes parámetros a usarse en el cálculo, separados por espacios y en el orden:

```
GA_POPULATION GA_ITERATION GA_CROSSOVER GA_MUTATION GA_IDDLE
GA_METH_ELITISM GA_ELITISM GA_SELECTION GA_CROSSMETH
```

donde:

GA_POPULATION	Tamaño de la población (número entero par)
GA_ITERATION	Número máximo de iteraciones de la optimización.
GA_CROSSOVER	Tasa de <i>crossover</i> (entre 0 y 1).
GA_MUTATION	Tasa de mutación (entre 0 y 1).
GA_IDDLE	Número máximo de iteraciones seguidas permitido sin que el valor óptimo obtenido para la función objetivo no mejore tras un número especificado de pasos.
GA_METH_ELITISM	Reemplazo condicional/incondicional/elitismo:
	none                      Reemplazo incondicional
	elitism                    Elitismo
	elicondi                  Reemplazo condicional
	elicondirepe              Reemplazo condicional sin repetición de genomas

GA_ELITISM	En caso de elitismo, este entero indica el número de individuos de la generación parental que pasan incondicionalmente a la generación siguiente.
GA_SELECTION	Métodos de selección (Tabla 8.1).
GA_CROSSMETH	Métodos de <i>crossover</i> (Tabla 8.2).

**Tabla 8.1.** Códigos para los métodos de selección.

1	<i>Roulette Wheel selection</i> (previo escalado de la función de <i>fitness</i> )
2	<i>Stochastic Remainder Selection Without Replacement</i> (previo escalado de la función de <i>fitness</i> )
3	<i>Roulette Wheel selection</i> (con ordenación de los valores de <i>fitness</i> )

**Tabla 8.2.** Códigos para los métodos de *crossover* y mutación

Codificación en bits	
1	<i>single-point crossover</i> con mutación aplicable a un único alelo
2	<i>two-point crossover</i> con mutación aplicable a un único alelo
3	<i>uniform crossover</i> con mutación aplicable a un único alelo
Codificación en enteros	
1	<i>single-point crossover</i> con mutación aplicable a todos los alelos
2	<i>two-point crossover</i> con mutación aplicable a todos los alelos
3	<i>uniform crossover</i> con mutación aplicable a todos los alelos
4	<i>single-point crossover</i> con mutación aplicable a un único alelo
5	<i>two-point crossover</i> con mutación aplicable a un único alelo
6	<i>uniform crossover</i> con mutación aplicable a un único alelo

En la Figura 8.2 se muestra un ejemplo para una selección *full array* por diversidad (\$SUBDISI) con el algoritmo de *Optimum Binning* (BIN), optimizando el recubrimiento de espacio (SPC)

```
$SUBDISI BIN 9 spc
30 20000 0.8 0.6 2000 elitism 3 3 1
```

**Figura 8.2.** Especificación de cálculo de un algoritmo genético.

## 8.2. Diseño de quimiotecas focalizadas

Tal y como se expone en el apartado 1.10.3, la forma más común de expresar la similitud de un conjunto de  $n$  compuestos en un espacio de descriptores multidimensional a un conjunto de  $F$  *leads* corresponde a la ecuación [1.114], definida como Similitud Mínima. Junto a esta, se incorporan en PRALINS las definiciones:

- Similitud Mínima (*Minimum Similarity*). Suma de distancias mínimas ( $d_{ij}$ ) de cada  $i$  compuesto en el conjunto a su *lead* más cercano, normalizada por el número de compuestos (ecuación [8.1] o [1.114]):

$$S(C) = \frac{1}{n} \sum_{i=1}^n \min_{j=1}^F (d_{ij}) \tag{8.1}$$

- Similitud Promedio (*Average Similarity*). Para cada  $i$  compuesto en el *array*, se calcula la suma normalizada sobre todas las distancias entre él y todos los *leads*. El índice de similitud corresponde al promedio para todas las moléculas en el conjunto (ecuación [8.2]):

$$S(C) = \frac{1}{n} \sum_{i=1}^n \frac{1}{F^2} \sum_{j=1}^F d_{ij} \quad [8.2]$$

- Similitud Máxima (*Maximum Similarity*). Como la similitud mínima, pero calculando la distancia de cada compuesto a su *lead* más distante (ecuación [8.3]):

$$S(C) = \frac{1}{n} \sum_{i=1}^n \max_{j=1}^F (d_{ij}) \quad [8.3]$$

- Similitud Centrada en un punto (*Centered point similarity*). Definida como el promedio de las distancias ( $d_{i,aver}$ ) entre cada compuesto y un *lead* virtual localizado en el centro de todos los *leads* en el espacio multidimensional (ecuación [8.4]):

$$S(C) = \frac{1}{n} \sum_{i=1}^n d_{i,aver} \quad [8.4]$$

Las distancias  $d_{ij}$ ,  $d_{i,aver}$  entre compuestos pueden calcularse con cualquiera de los coeficientes implementados previamente en PRALINS (distancia Euclídea, Manhattan, Tanimoto, coseno, Dice).

Estas cuatro definiciones de similitud están disponibles tanto para la selección en formato *cherry picking* como en formato *full array*. Además, para el caso de selecciones *cherry picking* se introducen otras cuatro alternativas posibles, no consideradas en el formato *full array* por su mayor requerimiento computacional respecto a las definiciones anteriores.

- Similitud distribuida en compuestos: El número total de compuestos se distribuye de forma que para cada *lead* se escogen el mismo número de compuestos más cercanos a cada uno de ellos. El resto de compuestos hasta completar el tamaño de selección requerido se escogen con i) el criterio de Similitud Promedio o ii) el criterio de Similitud Centrada en un punto.
- Se ordenan los compuestos de la quimioteca en función de su similitud centrada a un *lead* virtual y para cada *lead* se escoge el mismo número de compuestos más próximo a cada uno de ellos, en el orden calculado, hasta completar el tamaño de selección requerido.
- Se ordenan los compuestos de la quimioteca en función de su similitud promedio a todos los *leads* y para cada *lead* se escoge el mismo número de compuestos más próximo a cada uno de ellos, en el orden calculado, hasta completar el tamaño de selección requerido.

En una selección *cherry picking* simplemente se ordenan los compuestos en función de su similitud mediante el algoritmo *heapsort*<sup>303</sup>, escogiéndose aquellos en las primeras posiciones del *heap*.

Como se ha introducido, para las selecciones *full array*, el número de posibles selecciones de tamaño  $n$  en un conjunto  $N$  viene dado por la ecuación [8.5], donde  $i$  corresponde al número de *steps* de la quimioteca combinatoria y  $N_i$ ,  $n_i$  corresponden al número total de reactivos y de reactivos a seleccionar para el paso  $i$ .

$$\prod_{i=1}^{nsteps} \binom{N_i}{n_i} \quad [8.5]$$



Para escoger aquel conjunto que maximice la similitud, se implementan en PRALINS los métodos de optimización listados a continuación y presentados de manera general en el apartado 1.9. Los tres primeros de ellos son simplemente una translación de su versión original orientada a optimizar diversidad, aunque para facilitar la interpretación de los resultados mostrados en el apartado 8.2.2, se comentan sus especificaciones particulares, sobre todo en lo referente a los criterios de convergencia.

- Complete. Búsqueda completa. Evalúa de manera sistemática la similitud de todas las subquimiotecas posibles, por lo que solamente debe usarse en quimiotecas de tamaño reducido.
- Monte Carlo (MTC). Comenzando con una selección *full array* inicial, aleatoria o fijada por el usuario, en cada iteración se genera una nueva solución cambiando aleatoriamente un reactivo de un punto de sustitución también aleatorio. Si el índice de similitud mejora, la nueva solución reemplaza a la previa. De lo contrario, la nueva solución se acepta con una probabilidad determinada por el criterio de Metropolis (apartado 1.9.1). La constante de Boltzmann se calcula a partir de la diferencia promedio del valor de la función objetivo ( $\Delta f$ ) en aquellas transiciones de las 100 primeras en las que la función empeora. Se establece por convenio que a una temperatura de 300 K la probabilidad de aceptación sea del 20% (ecuación [8.6]). El algoritmo termina tras una serie de iteraciones definidas por el usuario.

$$k_B = \frac{\Delta f}{300 \cdot \ln 0.2} \quad [8.6]$$

- Simulated annealing (SA). Implementación análoga a la anterior, aunque en este caso la temperatura del criterio de aceptación no es fija, sino que disminuye desde un máximo que implica una fracción de aceptación de 0.9. Para este máximo se calcula la constante  $k_B$  mediante la expresión [8.7].  $k_B$  disminuye con una rampa de 0.99 por ciclo en que la nueva solución empeora y el algoritmo termina cuando la fracción de aceptación es inferior a 0.3.

$$k_B = \frac{\Delta f}{300 \cdot \ln 0.9} \quad [8.7]$$

- Algoritmos genéticos (GA). Siguiendo la implementación expuesta en el apartado 8.1. Para el estudio que prosigue, tras varias pruebas empíricas, se utilizan los siguientes parámetros: 30 individuos de tamaño de población, 20000 iteraciones hasta convergencia (2000 *idle*), el método *one-point crossover* con un tasa de *crossover* de 0.8 y una tasa de mutación de 0.6. Como método de selección se escoge el *Roulette Wheel selection* con ordenación de los valores de *fitness*. Para facilitar la convergencia, se aplica elitismo con una tasa del 10% de la población (3 individuos).
- Algoritmo Greedy (GREEDY). Se adopta como método de referencia el algoritmo propuesto por Agrafiotis y colaboradores.<sup>162</sup> El algoritmo comienza a partir de una *full array*  $n_1 \times n_2 \times \dots \times n_i$  seleccionado aleatoriamente a partir de una quimioteca  $N_1 \times N_2 \times \dots \times N_i$ . Entonces, secuencialmente recorre todos los puntos de sustitución  $R_1 \times R_2 \times \dots \times R_i$ , seleccionando los  $n_j$  reactivos para un *step* particular  $j$  que maximizan la función objetivo de similitud, mientras que se mantienen fijos el resto de reactivos fijados para el resto de puntos de diversidad. Por ejemplo, para  $j=1$ , los  $n_1$  reactivos se seleccionan generando un total de  $N_1$  arrays con configuración  $1 \times n_2 \times \dots \times n_i$  resultantes de la combinación de cada uno de los  $N_1$  candidatos disponibles para el punto de sustitución  $R_1$  y el resto de listas de reactivos ya seleccionados para los otros puntos de diversidad ( $R_2 \dots R_i$ ). Tras ordenar estos

$N_i$  arrays en orden de similitud decreciente, se seleccionan los mejores  $n_i$  arrays, que proporcionan los  $n_i$  reactivos para  $j=1$ .

Una vez que el proceso se ha repetido para todos los *steps*, se termina un ciclo y se compara la similitud del *full array* resultante con la de la solución anterior. Si la similitud mejora el algoritmo continua, de lo contrario termina. Se han implementado dos versiones que simulan el esquema de reemplazo de los algoritmos genéticos. En el primero, denominado *GREEDY-steady-state* (GREEDY-SS), tan pronto como se procesa un punto de variabilidad, los reactivos seleccionados reemplazan a aquellos ya presentes en la lista. En el denominado *GREEDY-generational* (GREEDY-GE) los reactivos seleccionados para cada punto de sustitución se insertan al final del ciclo.

- **Método Direct (DIRECT).** Se propone este nuevo método basándose en la idea de que en las quimiotecas combinatorias puede asumirse que los compuestos que presentan un *building block* común (reactivo) tendrán más probabilidad de ser más similares entre ellas que compuestos que no compartan dicho *building block*. Este método procede en tres etapas:
  - Selección del compuesto más próximo a un *lead* o conjunto de *leads*.
  - Deconvolución del compuesto seleccionado en sus reactivos constituyentes o *building blocks*, que son añadidos a la lista de reactivos seleccionados
  - Si el tamaño obtenido de selección para cada punto de diversidad no alcanza el tamaño requerido, el algoritmo vuelve al primer punto.

En realidad, el algoritmo procede seleccionando inicialmente un número suficiente de los compuestos más similares al *lead* o conjunto de *leads*. Esto coincide con una selección *cherry picking*, por lo que se contempla la posibilidad de que el criterio de similitud empleado en este punto corresponda a una de los ocho alternativas descritas anteriormente. Dado que éstas utilizan el algoritmo *heapsort* para ordenar la matriz de distancias, la implementación se realiza eficientemente, evitando una ordenación de todos los compuestos de la base de datos y limitando el tamaño resultante del *heap* de forma que asegure que la deconvolución de productos proveerá el número requerido de reactivos en cada punto de diversidad. En la Figura 8.3 se esquematiza el diagrama de flujo de dicho algoritmo.

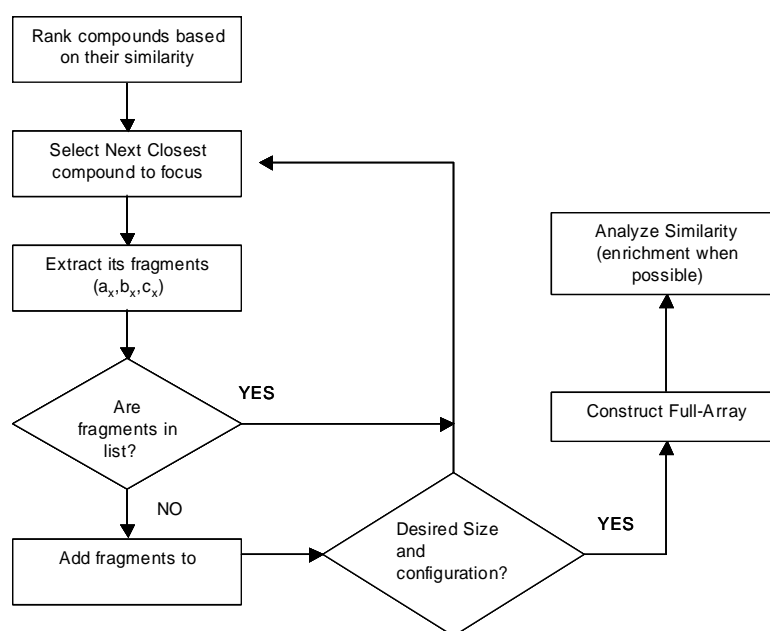


Figura 8.3. Diagrama de flujo del algoritmo Direct.

En todos los métodos se contempla la posibilidad de fijar los productos/reactivos requeridos. El concepto de grupo está disponible en todos los métodos excepto el método Direct. De nuevo, en todos ellos es posible iniciar la optimización a partir de una quimioteca propuesta por el usuario, exceptuando el método Direct (por su propia naturaleza determinista). Excepto la evaluación exhaustiva y el método Direct, el resto de métodos son no deterministas.

### 8.2.1. Instrucciones de cálculo en PRALINS del módulo de similitud

Las búsquedas de similitud se activan mediante la palabra clave \$FOCUS. A continuación se indica si se requiere o no la inclusión del compuesto(s) *lead(s)* en la selección final (YES/NON) así como el índice de los compuestos de la quimioteca considerados *leads* separados por comas.

Para las selecciones *cherry picking* se introduce la palabra clave \$METHFOCUS seguida del código numérico del criterio de similitud que se desea establecer, cuya correspondencia se lista en la Tabla 8.3.

**Tabla 8.3.** Correspondencia del valor de \$METHFOCUS con las definiciones de similitud.

Código numérico	Definición de Similitud utilizada en selecciones <i>cherry picking</i>
1	Focalización sobre un único <i>lead</i> *
2	Similitud Centrada en un punto ( <i>Centered point similarity</i> )
3	Similitud Promedio ( <i>Average Similarity</i> )
4	Similitud Mínima ( <i>Minimum Similarity</i> )
5	Similitud Máxima ( <i>Maximum Similarity</i> )
6	Similitud Centrada a un <i>lead</i> virtual + distribución en compuestos
7	Similitud Promedio + distribución en compuestos
8	Similitud distribuida en compuestos+criterio de Similitud Centrada en un punto
9	Similitud distribuida en compuestos+criterio de Similitud Promedio

\* esta opción se encuentra programada de manera independiente, ya que en el caso más común de focalizar a un único *lead* todas las definiciones de similitud coinciden.

Para las selecciones *full array* se introduce la línea con la palabra clave \$SUBMETHFOCUS seguida del indicativo del tipo de definición de similitud (Tabla 8.4) y del método de optimización a considerar (Tabla 8.5). Análogamente al módulo de diversidad, en el caso de escoger MTC se debe indicar en la línea siguiente la temperatura y el número de ciclos (ejemplo: T 300 MXC 1000). En el caso de escoger GA, los parámetros se indican en la línea siguiente tal y como se ha expuesto en el apartado 8.1.1. En el caso del método Direct, debe introducirse también la línea de especificación \$METHFOCUS que indique la definición de similitud a utilizarse en la selección *cherry picking* inicial (Tabla 8.3).

**Tabla 8.4.** Palabras clave para definición de similitud en selecciones *full array*.

Palabra clave	Definición de Similitud utilizada en selecciones <i>full array</i>
MINFOCUS	Similitud mínima ( <i>Minimum Similarity</i> )
AVGFOCUS	Similitud promedio ( <i>Average Similarity</i> )
MAXFOCUS	Similitud Máxima ( <i>Maximum Similarity</i> )
PFOCUS	Similitud centrada en un punto ( <i>Centered point similarity</i> )

**Tabla 8.5.** Algoritmos de optimización y código numérico a especificar en el fichero de entrada.

Código numérico	Método de optimización
1	Complete
3	MTC
4	SA
6	GREEDY-SS
7	GREEDY-GE
8	DIRECT
9	GA

En la Figura 8.4 se ejemplifican las especificaciones para la selección focalizada en diseño *cherry picking* (Figura 8.4.a) y *full array* con el método Direct (Figura 8.4.b) y el método GREEDY-SS (Figura 8.4.c).

<pre>\$FOCUS YES 8,14 \$METHFOCUS 4</pre>	<pre>\$FOCUS NON 13 \$SUBMETHFOCUS 8 \$METHFOCUS MINFOCUS 1</pre>	<pre>\$FOCUS YES 9,32 \$SUBMETHFOCUS AVGFOCUS 6</pre>
a	b	c

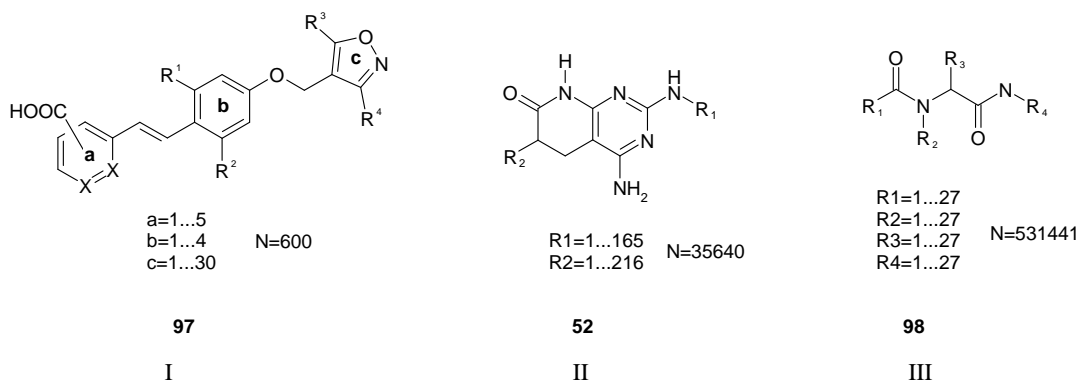
**Figura 8.4.** Especificaciones para el cálculo de varias selecciones focalizadas.

## 8.2.2. Análisis de los métodos de selección de quimiotecas *full array* focalizadas

La principal motivación de este análisis surge de la necesidad de testar el comportamiento del método Direct propuesto para la selección quimiotecas combinatorias focalizadas, así como comparar los diversos métodos de optimización implementados. Para ello, se escogen tres quimiotecas, cada una representativa de diferentes cardinalidades. El estudio se lleva a cabo mediante *scripts* programadas en perl.

### 8.2.2.1. Enumeración y descripción de las quimiotecas de estudio

- ***Quimioteca I.*** El primer caso de estudio es una quimioteca combinatoria compuesta por 600 oxazoles que ha sido previamente utilizada en la validación de las metodologías de selección por diversidad de PRALINS.<sup>166</sup> Desde un punto de vista combinatorio consta de tres puntos de variabilidad resultantes de la combinación de 5 carboxilatos, 4 halofenoles y 30 4-hidroximetilisoxazoles (Figura 8.5.a). Presenta la ventaja de que la actividad de los 600 compuestos como agonistas de FXR (receptor X farnesoide) ha sido testada, encontrándose que 36 compuestos muestran actividad equivalente a CDCA 50 $\mu$ M.<sup>461</sup> En su caracterización se emplearon un total de 100 descriptores del módulo Descriptor+ del programa Cerius2: fisicoquímicos, de forma, índices topoquímicos, índices topológicos basados en la teoría de la información, índices topoestructurales y contadores de subgrafos. La dimensionalidad de los datos se redujo mediante análisis de componentes principales a 7 componentes responsables del 90% de la varianza.
- ***Quimioteca II.*** Corresponde a la quimioteca BIB\_Amino de 35640 compuestos derivados de la reacción multicomponente de Victory expuesta en el capítulo 3 (Figura 8.5.b). En este estudio, se amplía el número de descriptores a un total de 84, calculados con MOE, que incluyen descriptores fisicoquímicos, índices topológicos basados en la teoría de la información, índices topológicos y descriptores espaciales. De nuevo, se aplica PCA para normalizar y decorrelacionar los datos, resultando en un espacio de 9 dimensiones que explican el 90% de la varianza total de los descriptores originales.
- ***Quimioteca III.*** El último caso de estudio corresponde a una quimioteca basada en la reacción de Ugi (Figura 8.5.c). Dicha quimioteca se deriva de la combinación de 27 ácidos (R<sup>1</sup>), 27 aminas (R<sup>2</sup>), 27 aldehídos (R<sup>3</sup>) y 27 isonitrilos (R<sup>4</sup>). Como se diseña con fines comparativos, estos reactivos se extraen aleatoriamente del catálogo Maybridge. La quimioteca final, compuesta por 531441 compuestos, se enumera con Cerius2 y se optimiza con MOE usando el *force field* MMFF94 con un gradiente de convergencia de 0.01 kcal/mol. Para su caracterización se emplean los mismos 84 descriptores usados para la quimioteca II. El espacio de propiedades final queda definido por 12 PCs que capturan el 90% de la varianza de los datos.



**Figura 8.5.** (I) Quimioteca I de oxazoles. (II) Quimioteca II de pirido[2,3-*d*]pirimidinas sustituidas BIB\_Amino. (III) Quimioteca III derivada de la reacción de Ugi.

Se desea destacar que debido a la naturaleza no aditiva de los descriptores empleados, puede asumirse que los resultados mostrados a continuación para el método Direct no están restringidos al uso de descriptores aditivos o semi-aditivos.

### 8.2.2.2. Quimiotecas combinatorias focalizadas entorno a un único lead

En primer lugar, se examina el comportamiento de los algoritmos de optimización DIRECT, SA, GA, GREEDY-GE y GREEDY-SS en el diseño de quimiotecas focalizadas entorno a un único lead.

Para cada quimioteca, se consideran *leads* de manera independiente un determinado número de compuestos escogidos aleatoriamente. Así, para la quimioteca I se realizan selecciones frente a 60 compuestos aleatorios (10% de la quimioteca), mientras que para las quimiotecas II y III se consideran 20 hipotéticos *leads*.

Para cada par particular quimioteca-lead se testa una amplia gama de tamaños de selección y configuraciones *full array*. Para la quimioteca I, el tamaño de selección máximo se establece en 600 compuestos, es decir, se permite seleccionar el total de la quimioteca. Para las otras dos quimiotecas se fija un valor máximo de 1000 (quimioteca II) y 10000 (quimioteca III) compuestos. Estos dos últimos valores de corte superan el valor de referencia, comúnmente aceptado para una selección representativa de la quimioteca total, de la raíz cuadrada del número total de compuestos de la quimioteca:  $\sqrt{35640} \approx 189$  y  $\sqrt{531441} \approx 729$ , respectivamente para cada quimioteca.

Respecto a las configuraciones requeridas, por simplicidad se opta por aquellos diseños  $n_1 \times n_2 \times \dots \times n_i$  directamente derivados de deconvolucionar repetidamente una selección *cherry picking* compuesta por  $x$  miembros. Es decir, para un lead específico de una de las quimiotecas, se ordenan sus  $x$  compuestos más cercanos en orden decreciente de similitud y se comienza deconvolucionando el primer compuesto de la lista. Obviamente, esto rinde un *array*  $1 \times 1 \dots \times 1$ . En la siguiente iteración, se aplica el mismo procedimiento al segundo mejor compuesto, identificándose una nueva configuración:  $1 \times 1 \dots \times 2$ , ...,  $2 \times 1 \dots \times 1$ ,  $2 \times 2 \dots \times 1$ , ..., ó  $2 \times 2 \dots \times 2$ . De aquí en adelante, los siguientes compuestos deconvolucionados pueden añadir un nuevo reactivo o *building block* a las listas de reactivos, proponiendo una configuración incrementada, o pueden estar formados por reactivos previamente seleccionados. Este proceso se repite hasta que el tamaño de la configuración obtenida excede el valor de corte impuesto (600, 1000 o 10000 compuestos). Nótese que de este modo los conjuntos de configuraciones a optimizar pueden diferir entre los distintos *leads* dentro de una misma quimioteca. Aproximadamente, un total de 30-40 *full arrays* son optimizados para cada lead hipotético de la quimioteca I, 90-100 para los de la quimioteca II y 20-30 para los de la quimioteca III.

A pesar del procedimiento impuesto, se remarca que PRALINS puede optimizar cualquier tipo de configuración requerido, tomando tantos compuestos de la lista como sean necesarios para alcanzar los tamaños de selección deseados para cada punto de variación de la quimioteca.

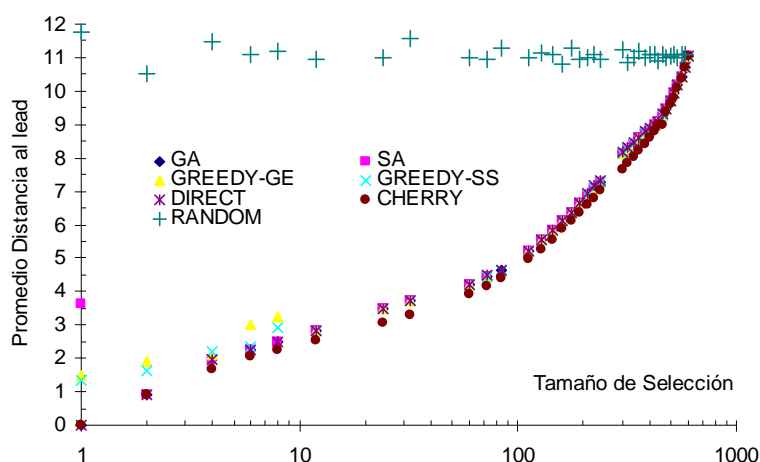
Estas selecciones se realizan también en formato *cherry picking* para los mismos tamaños de selección que las quimiotecas combinatorias.

El espacio de propiedades se define a través del conjunto ortogonal de componentes principales de cada quimioteca. Para todos los algoritmos, la función objetivo a optimizar corresponde a la definición de Similitud Mínima (ecuación [8.1]), aunque evidentemente en este caso ( $F=1$ ) cualquiera de las definiciones es equivalente. Como métrica se escoge la distancia Euclídea para medir las distancias entre compuestos, por lo que un menor valor del índice corresponde a una mayor similitud.

Para los algoritmos no deterministas (GA, SA, GREEDY-GE y GREEDY-SS) cada selección se repite 50 veces, cada vez comenzando con un *full array* aleatorio diferente. Únicamente el método Direct y el diseño *cherry picking* son totalmente deterministas.

Por una cuestión de espacio no se ilustran todos los diseños focalizados entorno a cada *lead*, sino que se muestra un único ejemplo para cada quimioteca para su inspección detallada y a continuación se presentarán valores estadísticos analizando las tendencias a lo largo del total de *leads* de cada quimioteca.

En la Figura 8.6 se ilustra el valor de similitud del conjunto de *full arrays* optimizados por los cinco métodos en torno a un *lead* hipotético de la quimioteca I. El tamaño de selección (eje de abcisas) se muestra en escala logarítmica para facilitar la visualización de los puntos. La selección *cherry picking* (*puntos marrones*) constituye el límite superior alcanzable de similitud (límite inferior en las figuras), ya que la restricción combinatoria incrementa el recubrimiento en las selecciones combinatorias focalizadas, por lo que supone una pérdida de similitud comparado con la selección *cherry picking*. Para los métodos no deterministas cada uno de los puntos mostrados corresponde al promedio de las 50 repeticiones.



**Figura 8.6.** Similitud a un compuesto *lead* para una serie de selecciones combinatorias focalizadas con tamaños de selección diferentes de la quimioteca I. En el eje de abcisas se representa en escala logarítmica el tamaño de selección frente al valor de similitud promedio de todas las repeticiones para un tamaño de selección concreto.

Para esta pequeña quimioteca ( $N=600$ ), se observa cómo las selecciones *full array* apenas se distancian de las selecciones *cherry picking*, a pesar de la restricción combinatoria. Para evaluar la importancia de estas diferencias, se considera como límite inferior el valor de similitud obtenido para una serie de selecciones combinatorias aleatorias (RANDOM).

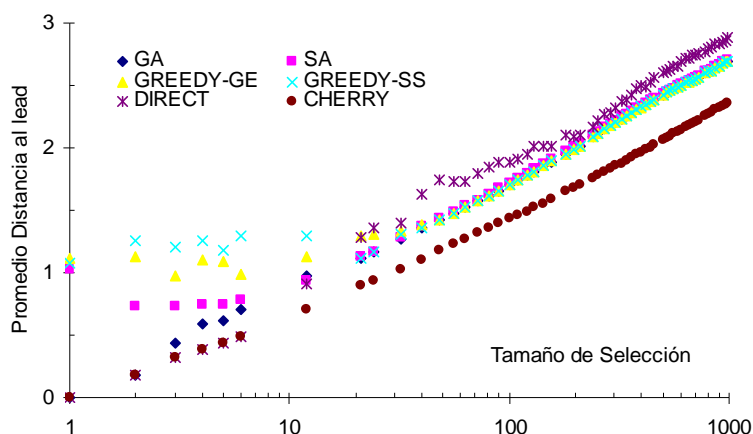
Por claridad, estos puntos (*cruces azules*) se representan exclusivamente para el caso de la quimioteca I (Figura 8.6). Cada punto aleatorio representa el promedio de 50 repeticiones. Así, al comparar las soluciones aleatorias no focalizadas, las diferencias entre las colecciones *cherry picking* y *full array* no son muy significativas. La tendencia relativamente constante de la selección aleatoria no debería confundir. Para estas selecciones, los puntos se encuentran uniformemente distribuidos en el espacio químico, por lo que la adición de nuevos compuestos apenas varía el valor promedio de similitud total. Por otra parte, para las selecciones focalizadas (*cherry picking* y *full array*), el valor de la función de similitud (distancia) se incrementa con el tamaño de selección, ya que los nuevos compuestos añadidos están más distantes del *lead* que los seleccionados previamente.

Para establecer y comparar la efectividad de los algoritmos de optimización, para cada tamaño de selección se calcula la diferencia entre el valor de similitud del correspondiente conjunto combinatorio identificado por un algoritmo concreto y el valor de similitud de la selección *cherry picking* del mismo tamaño de selección. A continuación, para cada algoritmo en particular se calcula el promedio de estas diferencias a lo largo del conjunto de tamaños de selección,  $V_{\text{ALGORITMO}}$  (ecuación [8.8]). Cuanto más se aproxime este valor a cero, mayor acuerdo existe entre las selecciones combinatorias y *cherry picking*.

$$V_{\text{ALGORITMO}} = \sum_{i=1}^{i=\text{tamaño seleccion}} S(C)_{\text{Full array ALGORITMO}} - S(C)_{\text{Cherry Picking}} \quad [8.8]$$

Sin embargo, debe considerarse el hecho de que la convergencia de los algoritmos no deterministas varía con el tamaño de selección. Así, para tamaños de selección superiores a 10 compuestos, los cuatro métodos (GA, SA, GREEDY-GE, GREEDY-SS) convergen prácticamente en la misma solución. Por debajo de este corte de 10 compuestos, mientras que las soluciones GA presentan una desviación estándar nula para las 50 repeticiones, SA converge ligeramente peor, particularmente en el caso de seleccionar un único compuesto. Por lo tanto, el promedio de las diferencias en similitud entre las selecciones combinatorias y *cherry picking* se calcula separadamente para las selecciones inferiores y superiores a 10 compuestos, observándose dos tendencias opuestas. Para las selecciones inferiores a 10 compuestos, los valores promedio para cada algoritmo ( $V_{\text{ALGORITMO}}$ ), ordenados por efectividad decreciente son: 0.181 (GA) = 0.181 (Direct) > 0.728 (GREEDY-SS) > 0.768 (GREEDY-GE) > 0.947 (SA). Así, GA y el método Direct son los óptimos. Por encima de este tamaño de selección, todos los algoritmos promedian igual (0.397), excepto el método Direct que ligeramente excede este valor (0.447).

Tocante a la quimioteca II, en la Figura 8.7 se ilustran los resultados obtenidos al muestrear diferentes tamaños de selección para un *lead* particular. De nuevo, todos los algoritmos rinden selecciones combinatorias muy próximas a las selecciones *cherry picking*. Los conjuntos *full array* focalizados se encuentran significativamente por debajo de la marca de similitud de las selecciones aleatorias, con un valor de  $4.711 \pm 0.053$  a lo largo de los diferentes tamaños de selección (por claridad no incluidas en la Figura 8.7).



**Figura 8.7.** Similitud a un compuesto *lead* para una serie de selecciones combinatorias focalizadas con tamaños de selección diferentes de la quimioteca II. En el eje de abcisas se representa en escala logarítmica el tamaño de selección frente al valor de similitud promedio de todas las repeticiones para un tamaño de selección concreto.

A partir de la convergencia de los algoritmos, se divide el rango de tamaños de selección en tres regiones. En la Figura 8.7 se observa cómo para tamaños pequeños de selección ( $<20$ ), los algoritmos GREEDY-GE, GREEDY-SS y SA quedan atrapados en un mínimo local, mientras que GA los sobrepasa. Para tamaños de selección inferiores a 50, el promedio de las desviaciones estándar para las 50 repeticiones de cada punto son: 0.155 (GA), 0.279 (SA), 0.399 (GREEDY-GE) y 0.356 (GREEDY-SS). Por encima de 50 compuestos, GREEDY-GE y GREEDY-SS conducen siempre a la misma solución (desviación estándar nula), mientras que la desviación estándar correspondiente para GA y SA es inferior que para tamaños de selección menores (0.00025 y 0.010, respectivamente). En este sentido, GA se comporta mejor que SA.

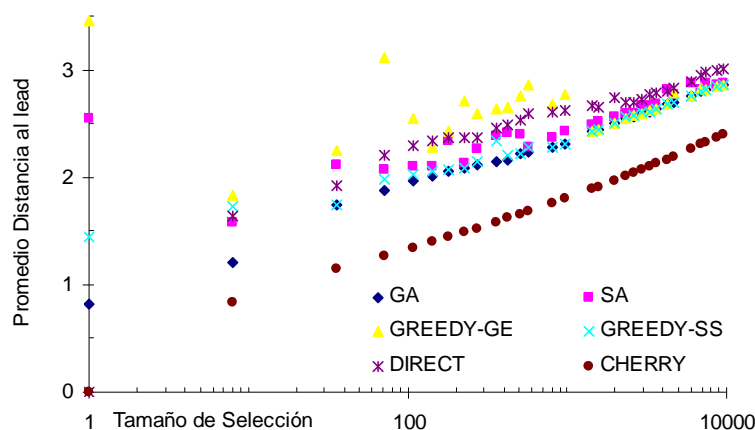
Los valores de  $V_{\text{ALGORITMO}}$  correspondientes a cada una de estas tres regiones se listan en la Tabla 8.6. Para tamaños de selección pequeños, el método Direct supera al resto de algoritmos, básicamente debido al problema de los mínimos locales (Figura 8.7). Para colecciones comprendidas entre 21-50 compuestos, GA es el mejor algoritmo, seguido de cerca por SA. A medida que el tamaño de selección se incrementa, las dos aproximaciones Greedy y GA rinden resultados idénticos (0.330), siendo SA menos efectivo. Para estos tamaños de selección, el método Direct rinde soluciones próximas a la óptima con un valor promedio de 0.495 (Tabla 8.6). Este valor es considerablemente inferior del  $V_{\text{ALGORITMO}}$  obtenido para las selecciones aleatorias referidas a las selecciones *cherry picking* de 2.810 (no mostrado en la Tabla 8.6).

**Tabla 8.6.** Valores  $V_{\text{ALGORITMO}}$  de las selecciones combinatorias focalizadas mostradas en la Figura 8.7 para la quimioteca II. Estos promedios se calculan para tres rangos de tamaños de selección.

Rango de selección	de tamaños de selección	GA	SA	GREEDY-SS	GREEDY-GE	Direct
n	[0-20]	0.149	0.428	0.673	0.782	0.074
n	[21-50]	0.242	0.252	0.288	0.248	0.475
n	[51-1000]	0.330	0.344	0.330	0.330	0.495

Finalmente, en la Figura 8.8 se representan los perfiles de similitud obtenidos al focalizar entorno a un *lead* de la quimioteca III. Las correspondientes selecciones aleatorias alcanzan un valor absoluto de  $4.447 \pm 0.335$  a lo largo del margen de tamaños de selección (por claridad no mostradas en la Figura 8.8).





**Figura 8.8.** Similitud a un compuesto *lead* para una serie de selecciones combinatorias focalizadas con tamaños de selección diferentes de la quimioteca III. En el eje de abcisas se representa en escala logarítmica el tamaño de selección frente al valor de similitud promedio de todas las repeticiones para un tamaño de selección concreto.

Para esta quimioteca, se escoge una cardinalidad de 200 compuestos para dividir el rango de tamaños de selección y calcular  $V_{\text{ALGORITMO}}$ . Para el *lead* específico mostrado en la Figura 8.8, por debajo de 200 compuestos, el ranking de los diferentes algoritmos de mejor a peor ( $V_{\text{ALGORITMO}}$ ) corresponde a: 0.608 (GA) > 0.763 (Direct) > 0.804 (GREEDY-SS) > 1.063 (SA) > 1.498 (GREEDY-GE).

Para  $n > 200$ , este ranking cambia: 0.525 (GA) > 0.542 (GREEDY-SS) > 0.613 (SA) > 0.711 (GREEDY-GE) > 0.739 (Direct). Las correspondientes selecciones aleatorias presentan valores de  $V_{\text{ALGORITMO}}$  de 3.317 ( $n < 200$ ) y 2.502 ( $n > 200$ ).

Tras inspeccionar un único *lead* para cada una de las tres quimiotecas, en la Tabla 8.7 se muestran las estadísticas a lo largo del conjunto completo de *leads* de cada quimioteca. Se combinan los resultados, expresados como la diferencia entre el valor de similitud del conjunto combinatorio y la selección *cherry picking* de todos los cálculos (número de *leads* × número de configuraciones analizadas para cada *lead* particular) para cada quimioteca. Estos promedios se calculan separadamente para cada uno de los rangos de tamaños de selección distinguidos anteriormente en cada caso.

**Tabla 8.7.** Valores promedio, expresados como la diferencia de similitud entre un conjunto combinatorio y una selección *cherry picking* de todos los cálculos (número de *leads* × número de configuraciones analizadas para un *lead* particular) realizados para una quimioteca particular.

	Quimioteca I		Quimioteca II			Quimioteca III	
	$n \leq 10$	$n > 10$	$n \leq 20$	$21 \leq n < 50$	$n > 50$	$n \leq 200$	$n > 200$
GA	0.122	0.332	0.107	0.146	0.256	0.474	0.584
SA	1.022	0.332	0.305	0.155	0.266	1.017	0.688
GREEDY-SS	0.585	0.332	0.399	0.230	0.259	0.741	0.592
GREEDY-GE	0.656	0.332	0.626	0.271	0.278	1.718	0.744
DIRECT	0.130	0.381	0.027	0.216	0.387	0.559	0.814

Respecto al ranking de los algoritmos en orden de efectividad decreciente se extraen las mismas conclusiones que las descritas para las Figuras 8.6-8.8. Para los tamaños de selección pequeños el método Direct identifica *full arrays* focalizados óptimos y a medida que el tamaño de selección se incrementa, su comportamiento es algo menos favorable, aunque sigue generando soluciones próximas a la óptima.

El comportamiento contrario se observa para los dos formatos del algoritmo GREEDY y SA, que conducen a selecciones más focalizadas a medida se incrementa el tamaño de selección, comparables a las obtenidas con GA. Particularmente, GREEDY-SS tiende a sobrepasar a SA y GREEDY-GE. El algoritmo genético es el que mejor se comporta en todos los casos, independientemente del tamaño de selección, y superando el problema de los mínimos locales encontrado para los otros algoritmos no deterministas.

Considerando la eficiencia de los algoritmos, en la Tabla 8.8 se muestra una lista comparativa del tiempo de cálculo requerido (uso de CPU en segundos) para realizar las diferentes selecciones para cada una de las tres quimiotecas. Todos los cálculos se ejecutan en un Pentium IV a 2.4 GHz en entorno Linux. El coste computacional asociado al método Direct es mucho más reducido, comparado con GA, y comparable al tiempo de las aproximaciones Greedy.

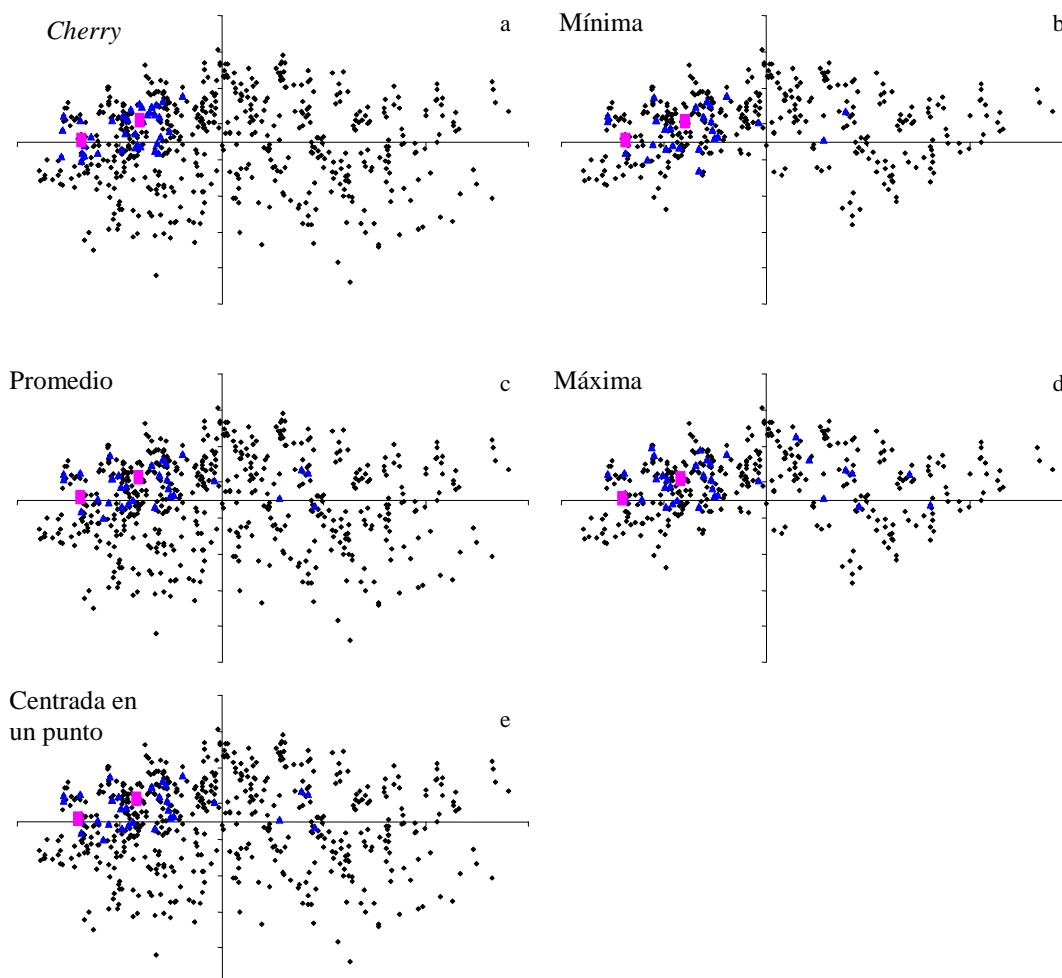
**Tabla 8.8.** Tiempo computacional requerido (uso de CPU en segundos) para la selección de diferentes colecciones para cada uno de los algoritmos de optimización descritos.

	Quimioteca I		Quimioteca II		Quimioteca III	
	n=24 4×2×3	n=192 4×4×12	n=182 13×14	n=988 26×38	n=810 5×9×3×6	n=4235 7×11×5×11
GA	1.18	5.89	9.32	44.13	32.98	193.31
SA	0.11	0.47	2.33	5.49	8.48	21.08
GREEDY-SS	0.05	0.05	1.83	1.84	6.16	6.39
GREEDY-GE	0.05	0.05	1.81	1.87	6.23	6.25
Direct	0.05	0.05	1.82	1.82	6.42	6.70

### 8.2.2.3. Quimiotecas combinatorias focalizadas entorno a varios *leads*

Se examina la efectividad de los algoritmos de optimización en el contexto del diseño de quimiotecas focalizadas entorno a varios compuestos *lead*,  $F > 1$  en las ecuaciones [8.1]-[8.4].

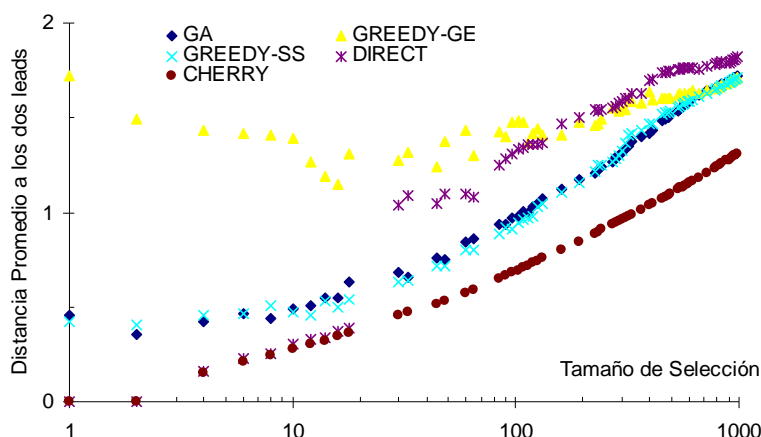
En este caso, las cuatro definiciones de similitud dejan de ser equivalentes. Por lo tanto, primero se analiza su comportamiento diferencial mediante inspección visual de la distribución de los compuestos en los conjuntos combinatorios optimizados utilizando las cuatro definiciones de similitud correspondientes. Para facilitar la representación, se escoge la quimioteca I ( $N=600$ ), con dos compuestos como *leads*. Se seleccionan *full arrays* de configuración  $2 \times 2 \times 10$ , optimizadas con GA. Como referencia, se toma una selección de 40 compuestos *cherry picking* ordenados según la ecuación [8.1]. Para simplificar la interpretación, la dimensionalidad de los 7 PCs se reduce de nuevo aplicando escalado multidimensional a la matriz de similitud entre pares de compuestos, donde las distancias se evalúan a partir de los 7 PCs (apartado 1.8). En la Figura 8.9 se ilustran los cinco diseños resultantes proyectados en dos dimensiones, que explican el 67% de la varianza de la matriz de distancias. Se observa cómo los cuatro conjuntos combinatorios (*triángulos azules* en las Figuras 8.9.b-8.9.e) cubren un área mayor que la selección *cherry picking* (*triángulos azules* en la Figura 8.9.a). En todas estas figuras, los cuadrados rosas corresponden a los dos *leads* hipotéticos.



**Figura 8.9.** Distribución de los 40 compuestos (*triángulos azules*) en los diferentes diseños focalizados a dos *leads* hipotéticos (*cuadrados rosas*). (a) selección *cherry picking* de los 40 primeros compuestos del ranking según ecuación [8.1]. (b) selección  $2 \times 2 \times 10$  optimizada con GA para Similitud Mínima ([8.1]). (c) selección  $2 \times 2 \times 10$  optimizada con GA para Similitud Promedio ([8.2]). (d) selección  $2 \times 2 \times 10$  optimizada con GA para Similitud Máxima ([8.3]). (e) selección  $2 \times 2 \times 10$  optimizada con GA para Similitud Centrada en un punto ([8.4]).

El diseño basado en la Similitud Máxima (ecuación [8.4]), que podría denominarse MiniMax, identifica colecciones distanciadas de los dos *leads*. Para esta configuración  $2 \times 2 \times 10$ , Similitud promedio (*Average Similarity*) y Similitud centrada en un punto (*Centered point similarity*) proponen diseños similares.

Tras inspeccionar las definiciones de similitud, se realizan selecciones frente a dos *leads* para cada una de las quimiotecas. Para cada par de *leads*, el conjunto de configuraciones y tamaños de selección a ser examinados se elige siguiendo el proceso descrito en el apartado anterior. Por simplicidad, únicamente se muestran los resultados obtenidos para la quimioteca II con los algoritmos GA, GREEDY-SS, GREEDY-GE y DIRECT (Figura 8.10). Como definición de similitud se utiliza el concepto de Similitud Mínima sobre el espacio de componentes principales y con métrica Euclídea. Estos resultados confirman el comportamiento apropiado de los métodos de focalización frente a selecciones aleatorias, que alcanzan unos valores de similitud de  $3.473 \pm 0.048$  (no mostrados en la Figura 8.10).



**Figura 8.10.** Similitud a **dos compuestos lead** para una serie de selecciones combinatorias focalizadas con tamaños de selección diferentes de la **quimioteca II**. En el eje de abscisas se representa en escala logarítmica el tamaño de selección frente al valor de similitud promedio de todas las repeticiones para un tamaño de selección concreto, calculado con ecuación [8.1].

Para tamaños de selección inferiores a 50 compuestos, el ranking de los métodos en orden de efectividad decreciente es: 0.176 (Direct) > 0.234 (GA) > 0.252 (GREEDY-SS) > 1.056 (GREEDY-GE) > 3.166 (Aleatoria). Por encima de este límite el ranking pasa a ser: 0.370 (GA) > 0.370 (GREEDY-SS) > 0.556 (GREEDY-GE) > 0.600 (Direct) > 2.466 (Aleatoria). Todos estos valores corresponden a  $V_{\text{ALGORITMO}}$ , es decir, el promedio de las diferencias entre la similitud de los *full arrays* y la similitud de la selección *cherry picking* a lo largo de los distintos tamaños de selección para un algoritmo determinado.

#### 8.2.2.4. Capacidad para identificar compuestos activos

En este punto se examina si la disminución en las medidas de similitud que conlleva el diseño combinatorio implica o no una disminución en el número de compuestos activos identificados por este diseño, comparada frente al caso de una selección *cherry picking*. Esto puede ser especialmente importante en el caso del método Direct, ya que a tamaños de selección moderados alcanza selecciones próximas a la óptima. Tal y como establece el Principio de Similitud, aquellas selecciones más focalizadas deberían idealmente alcanzar factores de enriquecimiento superiores. Sin embargo, esta premisa depende fuertemente de la capacidad del conjunto de descriptores utilizados para discriminar entre compuestos activos e inactivos. Además, incluso para descriptores que presentan un buen comportamiento de vecindad (*neighborhood behaviour*), el hecho de que un compuesto posea un alto índice de similitud frente a un compuesto activo no siempre asegura que este compuesto será activo. Por ejemplo, Martin y colaboradores<sup>24</sup> encontraron que para un valor de Tanimoto de 0.85 con los *fingerprints* Daylight, únicamente el 30% de los compuestos similares a un activo son ellos mismos activos.

La quimioteca I, compuesta por 600 oxazoles, es particularmente apropiada para un análisis retrospectivo de este tipo ya que i) ha sido totalmente testada frente a FXR, por lo que a diferencia de lo que se viene realizando en el trabajo, en este caso no es necesario presuponer la inactividad del *pool* de inactivos y ii) su naturaleza combinatoria.

Los *building blocks* contenidos en los 36 compuestos activos incluyen los 5 carboxilatos, los 4 halofenoles y 8 de los 30 4-hidroximetilisoxazoles. Así, como mínimo sería necesaria una quimioteca combinatoria de 160 compuestos ( $5 \times 4 \times 8$ ) para incluir los 36 compuestos activos. Previamente<sup>166</sup>, se demostró que los 7 componentes principales que describen esta quimioteca muestran un satisfactorio, aunque no perfecto, comportamiento de vecindad.

En este estudio, cada uno de los 36 oxazoles activos se considera separadamente como un compuesto *lead* para conducir búsquedas de similitud a tamaños de selección de 6, 30 y 60 compuestos, que representan el 1%, 5% y 10% del total de la quimioteca. Dichas selecciones se realizan en la forma de *sparse arrays* (*cherry picking*) y *full arrays*, estos últimos optimizados tanto por GA como por DIRECT. Para cada par particular *lead*-tamaño de selección, se testan todas las posibles configuraciones *full array* y aquella más focalizada (la de mayor índice de similitud) se elige para ser optimizada por GA y DIRECT. Como métrica se emplea la distancia Euclídea en el espacio descrito por los 7PCs derivados del set de 100 descriptores (apartado 8.2.2.1).

En la Tabla 8.9 se muestran los promedios de los índices de similitud obtenidos para los 36 compuestos activos a los tres porcentajes de base de datos cribada, así como el promedio de los factores de enriquecimiento (*ef*) encontrados.

**Tabla 8.9.** Promedio de los índices de similitud S(C) y los factores de enriquecimiento para las búsquedas de similitud focalizadas separadamente entorno a cada uno de los 36 compuestos activos a diferentes tamaños de selección (*n*): 1%, 5%, 10% de la quimioteca I. El espacio químico se define usando los 7 PCs derivados del set de 100 descriptores. La distancia Euclídea define las disimilitudes entre compuestos, por lo que un menor valor S(C) indica una mayor similitud.

	<i>n</i> = 6 (1%)			<i>n</i> = 30 (5%)			<i>n</i> = 60 (10%)		
	<i>Cherry</i>	GA	DIRECT	<i>Cherry</i>	GA	DIRECT	<i>Cherry</i>	GA	DIRECT
Promedio S(C)	2.25	2.70	2.77	4.03	4.72	4.89	5.07	5.71	5.81
Promedio <i>ef</i>	5.64	5.08	5.94	3.39	3.09	3.13	2.78	2.59	2.54

Se observa cómo la restricción combinatoria no reduce significativamente la tasa de identificación de compuestos activos. Incluso existen casos en los que una selección menos focalizada (por ejemplo con el método Direct en el primer 1% de base de datos) ligeramente supera los factores de enriquecimiento de selecciones más similares (el caso de GA y *cherry picking* en este 1%).

Para comprobar si esta inesperada tendencia es causada por el set de descriptores utilizados, se repiten las búsquedas de similitud usando otro set de descriptores bien establecido y validado, los descriptores MACCS.<sup>29</sup> Esto permite también el examen de la efectividad en términos de los valores de similitud encontrados con el método Direct bajo otros descriptores. Los 600 oxazoles se codifican con la versión empaquetada en bits del *fingerprint* MACC, implementada en MOE. Se define la similitud según la ecuación [8.1], usando el coeficiente de Tanimoto para medir la similitud entre pares de compuestos. En la Tabla 8.10 se listan los correspondientes promedios para los 36 activos de los índices de similitud y los factores de enriquecimiento.

**Tabla 8.10.** Promedio de los índices de similitud S(C) y los factores de enriquecimiento para las búsquedas de similitud focalizadas separadamente entorno a cada uno de los 36 compuestos activos a diferentes tamaños de selección ( $n$ ): 1%, 5%, 10% de la quimioteca I. El espacio químico se define usando las MACCS *keys*. El coeficiente de Tanimoto define las disimilitudes entre compuestos, por lo que un mayor valor S(C) indica una mayor similitud.

	$n = 6$ (1%)			$n = 30$ (5%)			$n = 60$ (10%)		
	<i>Cherry</i>	GA	DIRECT	<i>Cherry</i>	GA	DIRECT	<i>Cherry</i>	GA	DIRECT
Promedio S(C)	0.994	0.992	0.992	0.958	0.944	0.944	0.936	0.927	0.926
Promedio <i>ef</i>	6.02	6.79	5.71	4.40	4.12	3.83	3.07	3.04	3.16

De nuevo, el método Direct rinde valores de similitud próximos a los obtenidos con GA para la optimización de conjuntos combinatorios. A la vista de los factores de enriquecimiento, se observa cómo los descriptores MACCS se comportan mejor que los 100 descriptores usados anteriormente. Respecto a la influencia de la restricción combinatoria, se alcanzan las mismas conclusiones que en el caso anterior (Tabla 8.9). Por lo tanto, las diferencias en las medidas de similitud entre las selecciones *cherry picking* y *full array* no son tan relevantes cuando se analizan desde el punto de vista de factores de enriquecimiento, al menos para este caso de estudio.

A la vista de todos los resultados expuestos en los tres apartados anteriores, se puede concluir que la estrategia propuesta del método Direct, basado en una idea muy intuitiva, supone una aproximación válida para el diseño de quimiotecas focalizadas combinatorias en uno y varios *leads*. Tal y como sería esperable, a pequeños tamaños de selección el algoritmo rinde *full arrays* óptimos. A tamaños de selección mayores, rinde selecciones próximas a la óptima. Su comportamiento ha sido evaluado en diferentes condiciones: focalización frente a uno o varios *leads*, diversos tipos de representación química (variables reales y *fingerprints*) así como métricas de similitud (coeficiente de Tanimoto y distancia Euclídea). Desde un punto de vista de su capacidad para discriminar compuestos activos de inactivos, el método Direct está a la par de otros métodos como GA.

Dentro de los métodos de optimización aplicados a *full arrays*, GA es el que mejor opera tanto desde un punto de vista de su capacidad para maximizar la similitud como por su convergencia. Los otros métodos no deterministas muestran una peor convergencia a pequeños tamaños de selección, aunque se alcanzan buenos resultados a tamaños mayores. Respecto al coste computacional, GA es también el método más exigente, aunque para los tamaños de quimiotecas analizados, la velocidad computacional no es realmente importante.

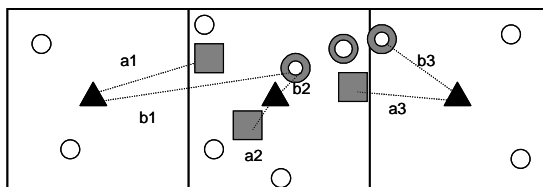
Así, para aquellas quimiotecas para las que la enumeración no suponga un paso limitante, el uso del algoritmo Direct representa un buen compromiso entre la habilidad para focalizar, la velocidad y su facilidad de implementación.

### 8.3. Criterios para evaluar la diversidad

En el estudio aquí realizado la diversidad de una selección se evalúa desde el punto de vista de su representatividad, es decir, el grado en que dicha selección se extiende en el rango de propiedades ofrecidas por la quimioteca total. Dentro de los criterios desarrollados para cuantificar esta diversidad destacan dos aplicados tradicionalmente: los métodos de partición o *cell-based* (apartado 1.10.2.3) y el *diversity integral criterion* (apartado 1.10.4). El primero de ellos se encuentra disponible en la versión original de PRALINS y ha sido aplicado en el capítulo 3, el segundo ha sido incorporado en el presente trabajo dentro de PRALINS.

Uno de los principales problemas asociados a los métodos de partición es la existencia de efectos frontera (*edge effects*) o artefactos del cluster.<sup>37,148,166,462</sup> En la Figura 8.11 se ejemplifica una situación hipotética de un artefacto del cluster. Se muestra una quimioteca virtual representada por los círculos pequeños en un espacio bidimensional, donde se seleccionan dos subconjuntos A (*cuadrados*) y B (*donuts*) de tres compuestos. Aunque visualmente el subconjunto B (*donuts*) está menos disperso que el subconjunto A (*cuadrados*), al dividir el espacio químico en tres celdas, la medida de recubrimiento de los subconjuntos A y B con la métrica *cell-based Fraction* (ecuación [1.103]) es de  $1/3$  y  $2/3$ , respectivamente. Por lo tanto, el subconjunto B sería cuantificado erróneamente como más diverso que A.

Una alternativa para superar este problema sería cuantificar distancias en lugar de evaluar recubrimiento en celdas. Si se colocan puntos de muestreo en el centro de cada celda o *bin* (*triángulos* en la Figura 8.11) y se evalúa la suma de las distancias mínimas de cada uno de los puntos a uno de los compuestos de cada subconjunto, el valor obtenido para el subconjunto A ( $a_1+a_2+a_3$ ) es inferior que el correspondiente para el subconjunto B ( $b_1+b_2+b_3$ ). Con ello, se consideraría correctamente que el subconjunto A muestrea más espacio químico que B.



**Figura 8.11.** Caso hipotético que representa la posibilidad de un artefacto del cluster. Subconjuntos A (*cuadrados*) y B (*donuts*) y puntos de muestreo colocados en el centro de cada *bin* (*triángulos*).

Este fenómeno del artefacto del cluster es particularmente importante en los diseños *full array*, ya que la restricción combinatoria previene un recubrimiento completo del espacio y favorece una distribución de los compuestos en el espacio más susceptible de sufrir estos artefactos. Además, es susceptible de ocurrir también en la evaluación de recubrimiento con los métodos de *clustering*, ya que no existe ninguna garantía de que las moléculas pertenecientes a un mismo cluster se encuentren más próximas entre ellas que con otras moléculas de otros clusters, tal y como Patterson y colaboradores han apuntado previamente.<sup>37</sup>

Basándose en la idea expuesta, se propone un nuevo criterio al que se ha bautizado como *cell-integral-diversity criterion*, abreviado C.I.D. Éste supone un compromiso entre los métodos de partición y el *diversity integral criterion* (en adelante referido como D.I). Como éste último, se basa en la medida de distancias entre puntos de muestreo (sondas) y compuestos en el espacio de descriptores, pero a diferencia de D.I, estas sondas no se posicionan aleatoriamente, sino que se colocan en el centro de las celdas o *bins* (Figura 8.11) ocupadas por las quimiotecas obtenidos por una técnica de partición como el algoritmo de *Optimum Binning*.

Así, para comparar dos quimiotecas que comparten el mismo espacio de descriptores, una vez particionado, la diversidad de cada quimioteca se cuantifica a partir de la suma de las distancias para cada uno de los puntos y su molécula más próxima. Aquella quimioteca con un menor valor, será más diversa.

C.I.D permite también solventar ciertas desventajas asociadas con el carácter aleatorio de los puntos utilizados por D.I, como son la reproducibilidad y el requerimiento de usar un elevado número de puntos.

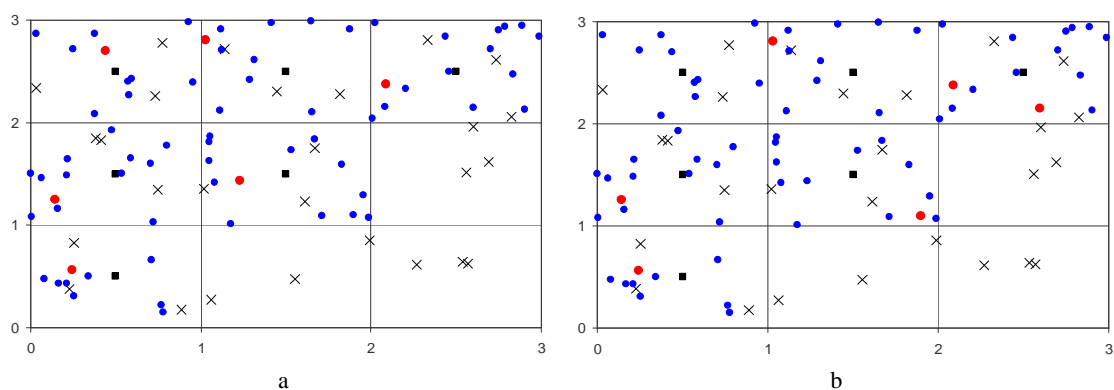
Aunque estos criterios C.I.D y D.I son muy similares en esencia, se debe tener en cuenta que el espacio real muestreado por ambos no es el mismo. En un espacio multidimensional definido por  $i$  propiedades ( $x_1...x_i$ ) los puntos aleatorios de D.I se posicionan en un hipercubo definido por las coordenadas  $[X_{Imin}...X_{Imax}]...[X_{imin}...X_{imax}]$ . Por lo tanto, puede existir, y normalmente es así, espacio químico no accesible a ninguno de los compuestos que está cubierto por los puntos de D.I, mientras que C.I.D únicamente tiene en cuenta las regiones hipercúbicas correspondientes al espacio químico ocupado. Esta diferencia tiene su importancia ya que en D.I los compuestos en los extremos del espacio de descriptores tendrán más importancia, favoreciendo el denominado efecto *edge design*<sup>43</sup>. En C.I.D no se favorece dicho efecto.

Este comportamiento se ilustra con el ejemplo de la Figura 8.12 para la evaluación de dos colecciones de 6 compuestos *A* y *B* (*puntos rojos* en las Figuras 8.12.a y 8.12.b, respectivamente) seleccionados de una quimioteca (*puntos azules*) en un espacio bidimensional. La colección *A*, inmersa en el espacio de la quimioteca, es claramente más representativa que la colección *B*, que difiere de *A* únicamente en dos compuestos situados en la periferia de los datos. Al particionar con *Optimum Binning* el espacio químico, el espacio real accesible comprende 6 de las 9 celdas totales, por lo que siguiendo el procedimiento C.I.D se colocan 6 puntos en el centro de ellas (*cuadrados negros*). Respecto a D.I, se colocan un total de 27 puntos aleatorios (*cruces negras*) uniformemente distribuidos en el espacio de propiedades, de manera que ocupan tres celdas sin compuestos. Se calcula C.I.D y D.I para cada colección de  $n$  miembros referido a la quimioteca total ( $N$ ) con la ecuación [8.9], donde  $p$  corresponde al número total de sondas y  $D_{ij}$ ,  $D_{ik}$  son las distancias Euclídeas entre una sonda y el compuesto más cercano de la colección ( $n$ ) y la quimioteca ( $N$ ), respectivamente.

$$D.I, C.I.D = \frac{\sum_{i=1}^p \min_{j \in n} D_{ij} - \sum_{i=1}^p \min_{k \in N} D_{ik}}{p} \quad [8.9]$$

En este ejemplo, se obtienen unos valores de 0.363 (C.I.D colección *A*), 0.527 (C.I.D colección *B*), 0.683 (D.I colección *A*) y 0.544 (D.I colección *B*). Así, según C.I.D, la colección *A* es más diversa que *B*, mientras que una tendencia opuesta se encuentra con *D.I*.





**Figura 8.12.** Influencia del espacio real muestreado por el *diversity integral criterion* (D.I) y *cell-integral-diversity criterion* (C.I.D). Se muestran dos subconjuntos seleccionados (*puntos rojos*) de una quimioteca (*puntos azules*): (a) subconjunto A, uniformemente distribuido y (b) subconjunto B con compuestos en la periferia de los datos. Los cuadrados negros representan las sondas de C.I.D y las cruces negras los puntos aleatorios de D.I.

En general, esta preferencia por los compuestos en los extremos del espacio de propiedades dependerá de la distribución particular de los compuestos en la quimioteca total, es decir, de la relación entre el espacio real accesible y el hipercubo definido para D.I.

### 8.3.1. Análisis de la eficacia y consistencia de los métodos de evaluación de diversidad

Como casos de estudio se consideran las tres quimiotecas presentadas en el apartado 8.2.2.1. La única diferencia respecto a lo allí expuesto es que para la quimioteca III basada en la reacción de Ugi, en lugar de considerar los 12 primeros componentes principales (90% de la varianza) se utilizan únicamente 7 (82% de la varianza) para definir el espacio químico, dado que es más conveniente usar los métodos de partición en espacios de dimensionalidad reducida (apartado 1.10.2.3).

Se compara el comportamiento de los tres criterios (*cell-based Fraction*, D.I y C.I.D) en la evaluación de diversidad de i) una serie de colecciones de diferente cardinalidad seleccionadas con el mismo método y ii) colecciones de la misma cardinalidad seleccionadas con métodos diferentes. La primera aproximación se aplica a las quimiotecas I y II y la segunda, a la quimioteca III. Para cada uno de los métodos de evaluación se analiza la influencia en su eficacia de los parámetros que usa (partición de referencia o el número de puntos a ser utilizados).

#### 8.3.1.1. Tamaños de selección analizados

Para las quimiotecas I y II se realizan un total de 21 selecciones *full array* con tamaños comprendidos en un rango centrado entorno a  $\sqrt{N}$ , siendo  $N$  el total de compuestos de la quimioteca:  $\sqrt{600} \approx 25$  (quimioteca I) y  $\sqrt{35640} \approx 189$  (quimioteca II). Para cada uno de estos tamaños de selección, se escoge aleatoriamente una configuración del *full array* ya que el objetivo de este trabajo no es determinar la mejor configuración “absoluta”, sino establecer una comparación metodológica. En la Tabla 8.11 se muestra el número de fragmentos a seleccionar para cada uno de los puntos de variación y el tamaño de selección total para cada quimioteca.

**Tabla 8.11.** Configuración de los 21 *full arrays* seleccionados para cada tamaño de selección en las quimiotecas I y II.

Quimioteca I				Quimioteca II		
Tamaño de selección	R1	R2	R3	Tamaño de Selección	R1	R2
15	5	1	3	160	10	16
16	2	2	4	165	11	15
17	1	1	17	170	17	10
18	2	3	3	175	25	7
19	1	1	19	180	12	15
20	2	2	5	185	37	5
21	3	1	7	190	19	10
22	2	1	11	200	10	20
23	1	1	23	205	5	41
24	2	2	6	210	14	15
25	5	1	5	215	43	5
26	2	1	13	220	11	20
27	3	3	3	225	15	15
28	2	2	7	230	23	10
29	1	1	29	235	5	47
30	3	2	5	240	20	12
32	2	2	8	245	35	7
33	3	1	11	250	10	25
34	2	1	17	255	15	17
35	5	1	7	260	20	13
36	3	2	6	265	53	5

Para la quimioteca III, se establece un tamaño de selección de 1296 compuestos, realizándose selecciones en formato *cherry picking (sparse)* y *full array* de configuración 6×6×6×6.

### 8.3.1.2. Métodos de selección aplicados

Para las quimiotecas I y II, las 21 selecciones de la Tabla 8.11 se optimizan bajo el criterio MaxMin promediado (ecuación [1.101]).

Para la quimioteca III, se aplican tres métodos (MaxMin promediado, *cell-based* con el algoritmo de *Optimum Binning* y *clustering* con el algoritmo *K-means*) para identificar los conjuntos de 1296 miembros tanto en formato *cherry picking* como *full array*. Además de estas seis colecciones, se lleva a cabo una selección aleatoria. Para los dos métodos de clasificación (*cell-based* y *clustering*) se optimiza la fracción de espacio cubierta (métrica *cell-based Fraction*, ecuación [1.103]). En el caso de las selecciones *cherry picking*, se escoge un compuesto aleatorio de cada *cluster* o *bin*.

Para las optimizaciones estocásticas (formatos *full array* y la versión promediada de MaxMin), se utiliza el algoritmo genético presentando con los mismos parámetros descritos en la sección 8.2. En todos los casos, se utiliza la distancia Euclídea para definir las distancias.

### 8.3.1.3. Condiciones de los métodos de evaluación

Una vez realizadas las selecciones para cada quimioteca, se determina su diversidad mediante los tres criterios (*cell-based Fraction*, D.I y C.I.D). En todos los casos, se define el espacio químico a partir de los mismos PCs empleados en la selección. Todas las selecciones y evaluaciones se realizan en PRALINS, con *scripts* en perl adicionales para automatizar el proceso.

- Método *cell-based*. Aplicando el algoritmo de *Optimum Binning*, el recubrimiento de cada colección se determina en función de la fracción de celdas ocupadas. Así, mayores valores indican mayor recubrimiento y diversidad.

De aquí en adelante, se referirá a esta métrica como SPC. Para demostrar el efecto del artefacto del *cluster* y estudiar el impacto de la partición de referencia escogida en el comportamiento del método, se realizan 6 particiones del espacio total en función del número de celdas/*bins* que ocupa la quimioteca total. En la Tabla 8.12 se lista el número de celdas ocupadas por cada una de las quimiotecas totales en cada uno de los 6 esquemas aplicados.

**Tabla 8.12.** 6 particiones de referencia utilizadas para evaluar la diversidad de las selecciones para cada una de las 3 quimiotecas.

Partición	Quimioteca I	Quimioteca II	Quimioteca III
	# Celdas ocupadas	# Celdas ocupadas	# Celdas ocupadas
P1	22	74	371
P2	40	76	459
P3	66	147	594
P4	112	213	646
P5	129	254	794
P6	200	442	833

- *Cell-integral-diversity criterion* (C.I.D). Para posicionar los puntos de muestreo se utilizan las seis particiones de la Tabla 8.12, donde el número de puntos totales equivale al número de celdas ocupadas. El valor obtenido para cada selección se refiere a la quimioteca total, siguiendo la ecuación [8.9]. A diferencia de SPC, un menor valor de C.I.D indica una mayor diversidad.

- *Diversity integral criterion* (D.I). Tradicionalmente este método se aplica en la comparación directa de dos selecciones de una quimioteca virtual. En este estudio, para comparar la diversidad de más de dos selecciones a la vez, el cálculo se realiza refiriendo la diversidad de cada una de las selecciones a la quimioteca total (ecuación [8.9]). El número de puntos aleatorios normalmente se ajusta a un valor lo suficientemente grande como para asegurar una buena distribución de los mismos. Inicialmente se escogen 1000 puntos para la quimioteca I y 10000 para las quimiotecas II y III. Además, para testar la influencia de este parámetro, se ejecutan seis cálculos con el mismo número de puntos que en el caso de C.I.D (Tabla 8.12). Debido a la naturaleza aleatoria del método, cada cálculo se repite 10 veces.

#### 8.3.1.4. Medida de la consistencia de los métodos de evaluación

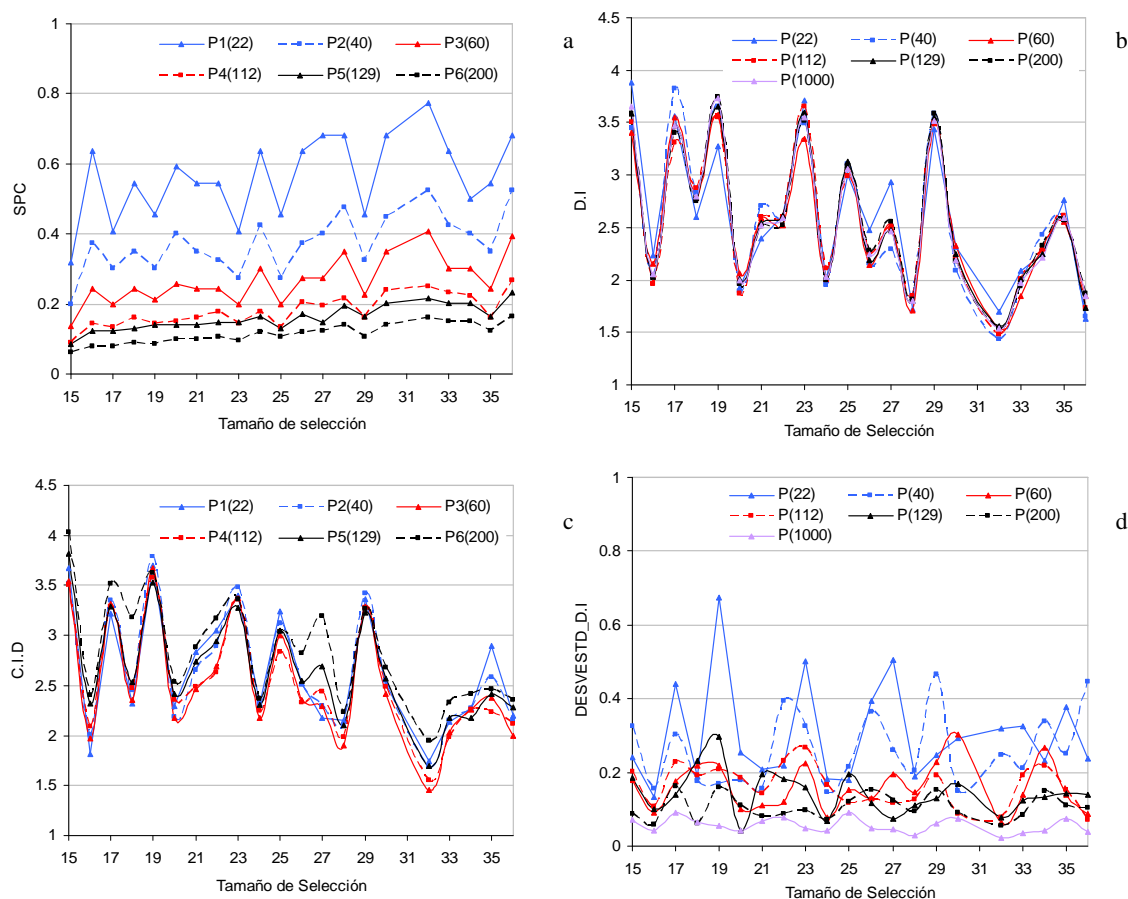
Para todos los experimentos de evaluación (definidos por un método concreto y sus parámetros) las tres series de selecciones correspondientes a cada quimioteca se ordenan según diversidad decreciente. Entonces se calcula el coeficiente de correlación de Spearman, ( $R_s$ , ecuación [7.2]) como una medida de la consistencia general en el ranking de las diferentes colecciones entre i) un método de evaluación bajo diferentes parámetros y ii) dos métodos de evaluación diferentes. Como se ha expuesto en el capítulo 7, este método se usa extensivamente en estudios de predicción de afinidad para correlacionar el ranking calculado con el ranking experimental de las moléculas candidatas. En este estudio, no existe un ranking predecible de las selecciones que deba reproducirse. Por ello, aquí este coeficiente es meramente usado para analizar la influencia de los diferentes parámetros en un método de evaluación particular (variación interna) y para establecer una comparación de los métodos.

#### 8.3.1.5. Resultados para las colecciones de diferente cardinalidad (quimiotecas I y II)

En las Figuras 8.13 y 8.14 se muestran las evaluaciones de diversidad de los tres métodos: SPC (a), D.I (b) y C.I.D (c) para las quimiotecas I y II, respectivamente. En los gráficos SPC y C.I.D, cada una de las líneas corresponde a cada uno de los esquemas de *binning* considerados como partición de referencia (P1...P6, Tabla 8.12).

En el caso de D.I, las seis líneas P1-P6 corresponden a las seis evaluaciones con un número de puntos aleatorios equivalente al número de celdas ocupadas en las seis particiones. La séptima línea adicional pertenece al caso de aplicar un total de 1000 (quimioteca I) o 10000 (quimioteca II) puntos. Estos valores de D.I son el promedio de las 10 repeticiones, cuya desviación estándar se muestra en las Figuras 8.13.d y 8.14.d.

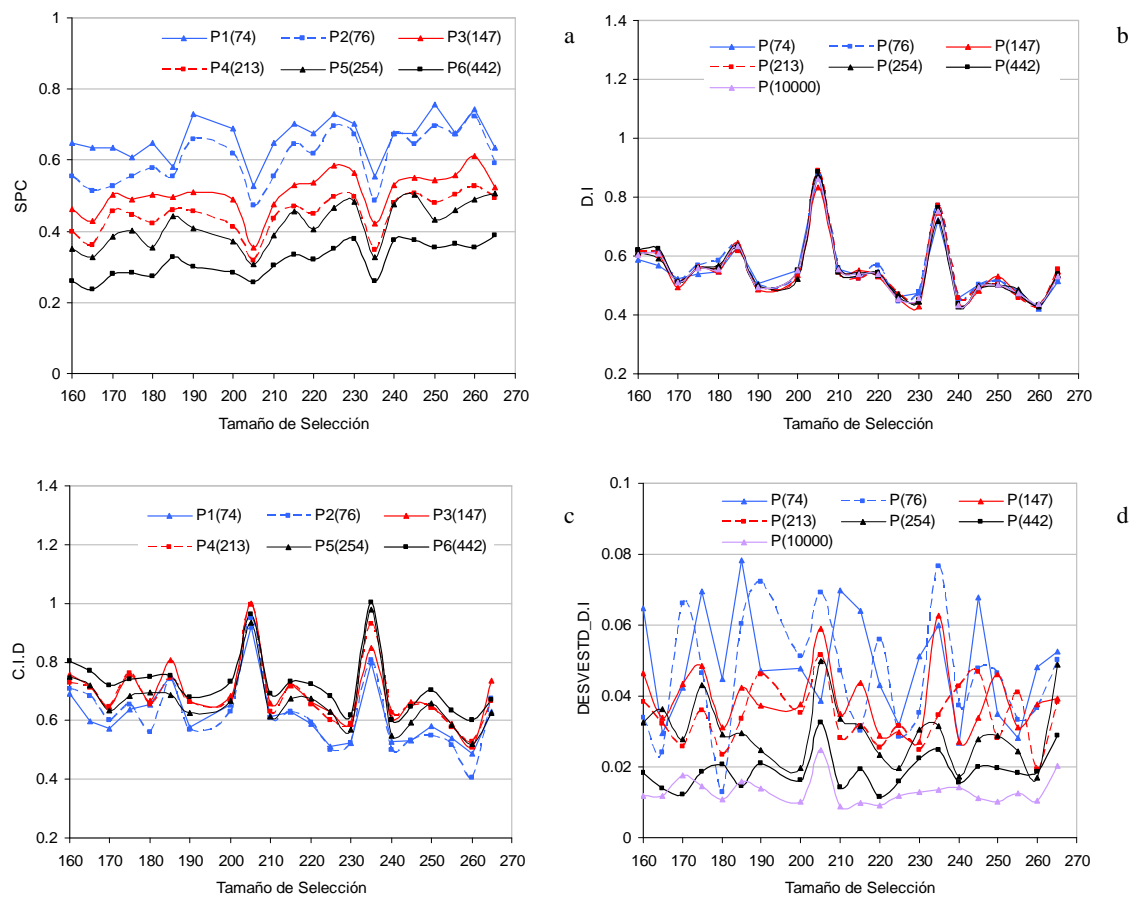
Como se ha mencionado, no existe un ranking óptimo predecible para las 21 selecciones en orden diversidad decreciente conocido *a priori* para ser considerado como referencia. Aunque en principio se esperaría que el recubrimiento se incremente con el número de compuestos seleccionados, esto es exclusivamente cierto en los casos en que se comparan selecciones con grandes diferencias de cardinalidad. En este caso, el paso de variación entre estas 21 selecciones no es significativo: 1 y 5 compuestos para la quimioteca I y II, respectivamente. Por ejemplo, para la quimioteca I se observa un ejemplo en el que la selección de 29 compuestos cubre menos espacio que la de 28, evaluada por cualquiera de los tres criterios (Figuras 8.13.a, b y c). Sin embargo, en promedio puede observarse una tendencia creciente de diversidad (al trasladarse de 15 a 36 compuestos en la quimioteca I).



**Figura 8.13.** Medidas de diversidad con los criterios (a) SPC, (b) D.I (valor promedio) y (c) C.I.D para cada una de las 21 selecciones de diferente cardinalidad realizadas para la **quimioteca I**. (d) Desviación estándar de las 10 repeticiones de D.I.

Se analiza en primer lugar la variación interna de las medidas de diversidad con el criterio SPC (Figuras 8.13.a y 8.14.a) bajo diferentes particiones. Conforme el número de *bins* de la partición de referencia se reduce, más probable resulta para una determinada selección alcanzar un mayor recubrimiento.

Por lo tanto, no sorprende que cada una de las 21 selecciones de las quimiotecas I y II devuelva mayor valor absoluto al ser evaluada bajo P1 que bajo P6. Sin embargo, el ranking de las 21 selecciones en orden de diversidad decreciente no siempre se conserva a lo largo de las seis particiones. Por ejemplo, considerando la quimioteca I, en la Figura 8.13.a, la selección de 35 miembros se sitúa por encima de la de 34 miembros al ser evaluada frente a P1, mientras que para el resto de particiones de referencia (P2-P6) se encuentra un orden opuesto. Una situación similar se observa para la quimioteca II (Figura 8.14.a): bajo P1, la colección de 250 miembros es más diversa que la de 245, invirtiéndose este ranking bajo P5. Para cuantificar el grado de consistencia, se calcula  $R_s$  para las 15 combinaciones posibles de pares de listas ordenadas de las 21 selecciones establecidas por cada uno de los 6 esquemas de *binning* ( $6 \times 5/2$ ). Estos resultados se muestran en las Tablas 8.13 y 8.14 para las quimiotecas I y II, respectivamente (primera-quinta columna). Aunque existe una correlación positiva, la partición de referencia presenta un elevado impacto en el ranking de las selecciones, con  $R_s$  comprendido entre 0.534 y 0.949 (quimioteca I) y entre 0.464 y 0.953 (quimioteca II).



**Figura 8.14.** Medidas de diversidad con los criterios (a) SPC, (b) D.I (valor promedio) y (c) C.I.D para cada una de las 21 selecciones de diferente cardinalidad realizadas para la **quimioteca II**. (d) Desviación estándar de las 10 repeticiones de D.I.

**Tabla 8.13.** Coeficientes de correlación de Spearman para el ranking de las 21 selecciones de la quimioteca I por las seis medidas de diversidad con SPC y el criterio D.I con 1000 puntos.

	SPC (P1)	SPC (P2)	SPC (P3)	SPC (P4)	SPC (P5)	D.I (P=1000)
SPC (P1)						0.842
SPC (P2)	0.922					0.944
SPC (P3)	0.845	0.949				0.918
SPC (P4)	0.590	0.673	0.730			0.753
SPC (P5)	0.534	0.645	0.657	0.894		0.641
SPC (P6)	0.705	0.801	0.861	0.677	0.653	0.734

**Tabla 8.14.** Coeficientes de correlación de Spearman para el ranking de las 21 selecciones de la quimioteca II por las seis medidas de diversidad con SPC y el criterio D.I con 10000 puntos.

	SPC (P1)	SPC (P2)	SPC (P3)	SPC (P4)	SPC (P5)	D.I (P=10000)
SPC (P1)						0.731
SPC (P2)	0.887					0.877
SPC (P3)	0.722	0.887				0.882
SPC (P4)	0.539	0.791	0.927			0.861
SPC (P5)	0.477	0.743	0.844	0.947		0.795
SPC (P6)	0.464	0.739	0.809	0.894	0.953	0.775

Respecto a las medidas D.I, puede observarse en las Figuras 8.13.b y 8.14.b que estas evaluaciones son prácticamente independientes del número de puntos aleatorios aplicados. De hecho, los valores de  $R_s$  de las 21 combinaciones de pares de las 7 listas ( $7*6/2$ ) promedian 0.967 (quimioteca I) y 0.969 (quimioteca II). Se remarca que los valores mostrados en estas figuras corresponden al promedio de 10 repeticiones. Si se ejecuta una única evaluación, se pierde un cierto grado de acuerdo entre las siete curvas y  $R_s$  decrece. En el caso de la quimioteca II, el valor promedio de  $R_s$  para las 21 combinaciones de ranking varía desde 0.969 (10 repeticiones) a 0.815 (1 repetición). Evidentemente, la magnitud de la desviación estándar obtenida depende tanto del número de sondas (a mayor número menor desviación estándar) como de la quimioteca particular bajo estudio (compárense las Figuras 8.13.d y 8.14.d).

En lo tocante a los gráficos de C.I.D, los perfiles de diversidad obtenidos son muy similares a los obtenidos para el índice D.I. En las Tablas 8.15 y 8.16 se listan los coeficientes  $R_s$  para las 15 combinaciones de pares de rankings obtenidos para las seis particiones de referencia (primera-quinta columna), para las quimiotecas I y II, respectivamente. Para ambas quimiotecas, las seis listas de selecciones ordenadas presentan un buen acuerdo, con  $R_s$  comprendido entre 0.821 y 0.992 (quimioteca I) y entre 0.812 y 0.984 (quimioteca II). Comparado con los valores de  $R_s$  obtenidos para el método *cell-based*, la influencia de la partición se reduce considerablemente con C.I.D. Esta diferencia puede ser atribuida a la eliminación del artefacto del *cluster* (Figura 8.11) por parte de C.I.D, ya que el resto de factores como la distribución intrínseca de los compuestos en los *bins* es equivalente en ambos casos.

**Tabla 8.15.** Coeficientes de correlación de Spearman para el ranking de las 21 selecciones de la quimioteca I por las seis medidas de diversidad con C.I.D y el criterio D.I con 1000 puntos.

	C.I.D (P1)	C.I.D (P2)	C.I.D (P3)	C.I.D (P4)	C.I.D (P5)	D.I (P=1000)
C.I.D (P1)						0.897
C.I.D (P2)	0.981					0.936
C.I.D (P3)	0.961	0.992				0.945
C.I.D (P4)	0.901	0.935	0.955			0.914
C.I.D (P5)	0.887	0.925	0.945	0.969		0.916
C.I.D (P6)	0.821	0.879	0.906	0.948	0.958	0.938

**Tabla 8.16.** Coeficientes de correlación de Spearman para el ranking de las 21 selecciones de la quimioteca II por las seis medidas de diversidad con C.I.D y el criterio D.I con 10000 puntos.

	C.I.D (P1)	C.I.D (P2)	C.I.D (P3)	C.I.D (P4)	C.I.D (P5)	D.I (P=10000)
C.I.D (P1)						0.926
C.I.D (P2)	0.903					0.923
C.I.D (P3)	0.899	0.927				0.868
C.I.D (P4)	0.892	0.917	0.983			0.869
C.I.D (P5)	0.847	0.812	0.818	0.855		0.904
C.I.D (P6)	0.873	0.843	0.826	0.862	0.984	0.935

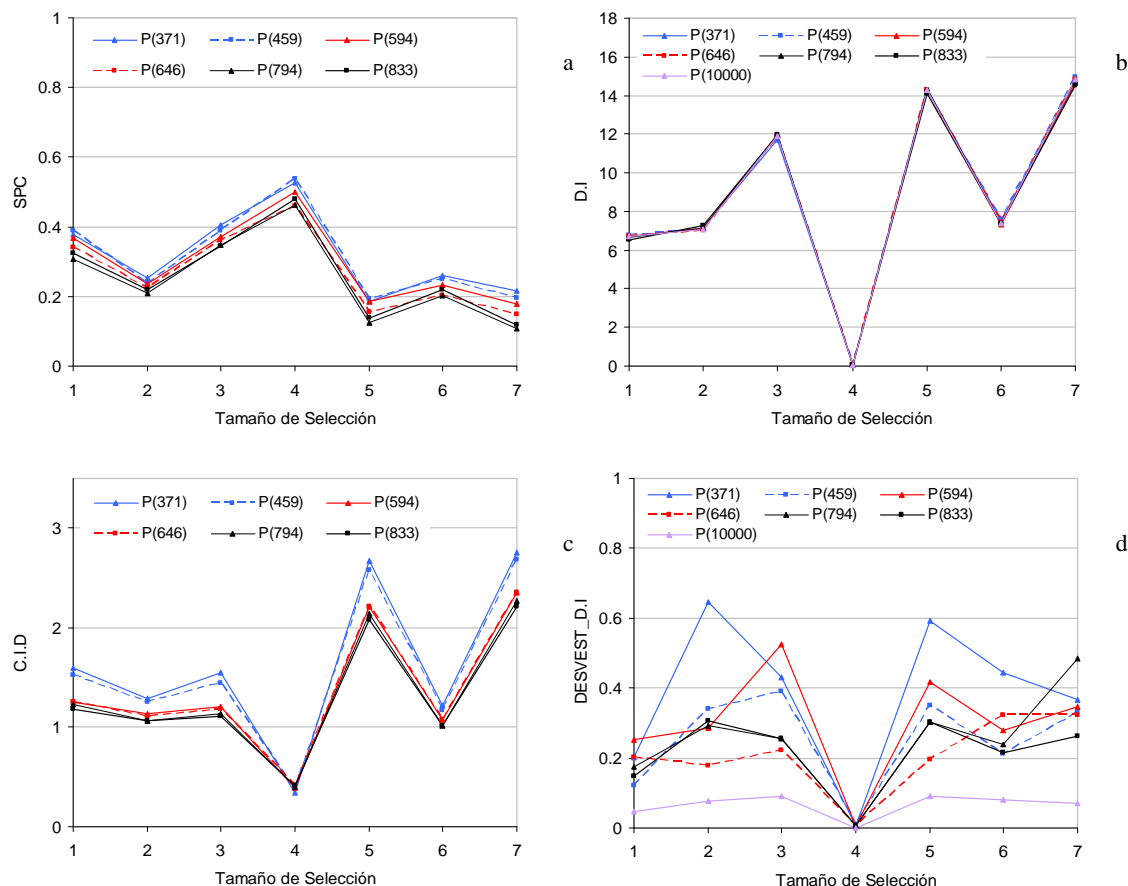
Tras juzgar la variación interna de cada método en función de sus diferentes condiciones intrínsecas de evaluación (particiones de referencia y número de puntos), se compara el grado de acuerdo entre los tres índices de diversidad.

Dada la coherencia mostrada a través de las 7 determinaciones, se escoge el ranking de las 21 colecciones obtenidas por el índice D.I con 1000 puntos (quimioteca I) o 10000 puntos (quimioteca II) como referencia. En la última columna de las Tablas 8.13-8.16 (etiquetadas "D.I") se resumen los correspondientes coeficientes de correlación entre los *full arrays* ordenados por el criterio D.I ( $p=1000$  o  $p=10000$ ) y cada una de las seis determinaciones realizadas con el criterio SPC (Tablas 8.13 y 8.14) y C.I.D (Tablas 8.15 y 8.16). Para las dos quimiotecas, D.I correlaciona mejor con el ranking de C.I.D que con el de SPC. Para la quimioteca I, el promedio de  $R_s$  para los pares D.I-C.I.D y D.I-SPC es de 0.924 y 0.805, respectivamente. Para la quimioteca II, estos valores son 0.904 (D.I-C.I.D) y 0.820 (D.I-SPC). Así, para las dos quimiotecas, el promedio de los coeficientes de correlación de los pares D.I-C.I.D es muy similar a los valores obtenidos para la variación interna del método D.I. Esta pequeña diferencia puede atribuirse a las diferencias en el espacio químico muestreado por uno u otro método (Figura 8.12).

### 8.3.1.6. Resultados de colecciones seleccionadas con distintos métodos (quimioteca III)

En la Figura 8.15 se muestran los tres índices de diversidad para las siete selecciones diversas de la quimioteca III (apartado 8.3.1.2). El ranking de los métodos bajo los criterios C.I.D y D.I permanece constante cuando se alteran las condiciones de evaluación, rindiendo por lo tanto valores de  $R_s=1$ . Respecto al criterio SPC, la variación de la partición de referencia identifica dos ordenaciones de los métodos casi idénticas que solo difieren en la posición en el ranking entre la selección *full array* con *K-means* y la selección aleatoria.

En la Tabla 8.17 se resumen los cuatro posibles rankings de las siete selecciones encontradas por los tres índices de similitud. Dentro de cada lista, los métodos aparecen en orden de diversidad decreciente. La selección *sparse* con *Optimum Binning* es claramente identificada como la más diversa por los tres índices. La selección *full array* con *K-means* es casi indistinguible de la selección aleatoria por cualquiera de los tres métodos (Figura 8.15.a-8.15-c), por lo que su intercambio de rango dependiendo de la partición de referencia del criterio SPC no supone una desventaja relevante. Así, los tres índices coinciden en los casos extremos de máxima (*cherry picking* con *Optimum Binning*) y mínima (*full array* con *K-means*) diversidad.



**Figura 8.15.** Medidas de diversidad con los criterios (a) SPC, (b) D.I (valor promedio) y (c) C.I.D para cada una de las 7 selecciones realizadas para la **quimioteca III** con diferentes métodos y diseños. (d) Desviación estándar de las 10 repeticiones de D.I. En el eje de abscisas se presenta el índice que identifica a cada colección. 1: MaxMin *full array*; 2: MaxMin *sparse array*; 3: *Optimum Binning full array*; 4: *Optimum Binning sparse array*; 5: *K-means full array*; 6: *K-means sparse array*; 7: Aleatoria.

**Tabla 8.17.** Ranking de las siete colecciones seleccionadas por diferentes metodologías para la **quimioteca III** en orden de diversidad decreciente calculada por los tres criterios: C.I.D, D.I y SPC.

Criterio de diversidad	C.I.D	D.I	SPC	
	P1-P6	Todos los sets de puntos	P1-P2	P3-P6
Rango 1	<i>Binning Sparse</i>	<i>Binning Sparse</i>	<i>Binning Sparse</i>	<i>Binning Sparse</i>
Rango 2	<i>K-means Sparse</i>	MaxMin <i>Full</i>	<i>Binning Full</i>	<i>Binning Full</i>
Rango 3	MaxMin <i>Sparse</i>	MaxMin <i>Sparse</i>	MaxMin <i>Full</i>	MaxMin <i>Full</i>
Rango 4	<i>Binning Full</i>	<i>K-means Sparse</i>	MaxMin <i>Sparse</i>	MaxMin <i>Sparse</i>
Rango 5	MaxMin <i>Full</i>	<i>Binning Full</i>	<i>K-means Sparse</i>	<i>K-means Sparse</i>
Rango 6	<i>K-means Full</i>	<i>K-means Full</i>	Aleatoria	<i>K-means Full</i>
Rango 7	Aleatoria	Aleatoria	<i>K-means Full</i>	Aleatoria

Es remarcable que el método *cell-based* prioriza aquellas selecciones realizadas con el algoritmo de *Binning*, ya que tanto el criterio de selección y evaluación coinciden. Los otros dos criterios, que cuantifican distancias en lugar de ocupación de celdas, otorgan un menor valor de diversidad a la selección *full array* con *Optimum Binning* frente a otras selecciones basadas en



distancia, incluyendo la selección *full array* con MaxMin así como las selecciones basadas en *clustering*, representadas por el *sparse array* identificado con *K-Means*.

Analizando las selecciones MaxMin, se observa que algunos compuestos situados en los extremos del espacio de descriptores son los responsables del valor de D.I calculado para estas colecciones, siguiendo la idea ilustrada en la Figura 8.12. Por lo tanto, estas colecciones aparecen en posiciones preferentes del ranking establecido por D.I. Sin embargo, debería considerarse que las evaluaciones D.I de la selección *cherry picking* con *K-means*, *sparse* con MaxMin y el *full array* con MaxMin son muy similares, con desviaciones estándar solapadas (Figuras 8.15.b y 8.15.d). Lo mismo sucede con el criterio C.I.D (Figura 8.15.c). Sin embargo, en este caso, es destacable que el método sitúa las selecciones *cherry picking (sparse)* por encima de los correspondientes diseños *full array*, una tendencia esperable ya que la restricción combinatoria implica una limitación para conseguir buen recubrimiento, lo cual no se observa con las otras métricas de evaluación.

### 8.3.1.7. Cell-integral-diversity criterion en el diseño de quimiotecas

Tras analizar la capacidad del *cell-integral-diversity criterion* para evaluar y establecer la diversidad, se examina el uso de C.I.D como métrica/función a optimizar en el contexto del diseño de quimiotecas. Particularmente, en el caso del diseño *full array* puede ser muy útil ya que este formato agrupa a los compuestos en el espacio de manera que los artefactos del *cluster* con los métodos *cell-based* son frecuentes.

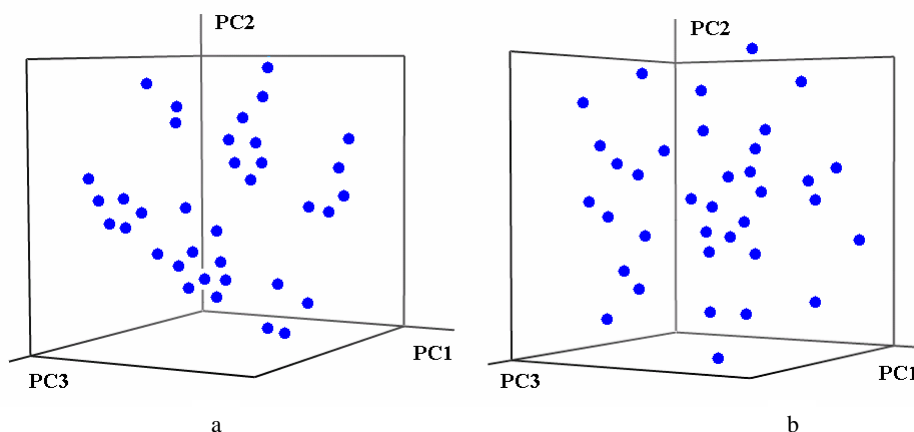
De forma resumida, su esquema de trabajo para las selecciones *full array* consiste en una partición *Optimum Binning* del espacio seguida de la colocación de las sondas en el centro de los *bins* ocupados y finalmente minimización de la función objetivo dada en la ecuación [8.10] por el algoritmo genético descrito.

$$C.D.I = \frac{\sum_{i=1}^p \min_{j \in n} D_{ij}}{p} \quad [8.10]$$

donde  $p$  corresponde al número de puntos situados en el centro de cada *bin* (número de *bins* ocupados).

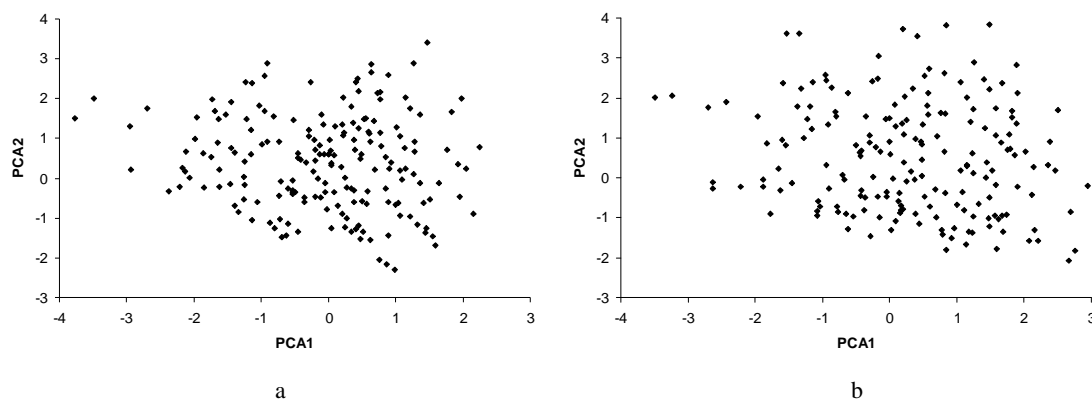
Se realizan tres selecciones, una para cada una de las quimiotecas (I, II y III), dirigidas a identificar la mejor quimioteca combinatoria  $3 \times 2 \times 6$ ,  $19 \times 10$  y  $6 \times 6 \times 6 \times 6$  utilizando la función C.I.D (ecuación [8.10]). Con fines comparativos, se optimizan las mismas selecciones usando la función *cell-based Fraction* (SPC). Para la quimioteca I, el algoritmo de partición genera 22 celdas ocupadas, mientras que 147 y 1209 *bins* ocupados se obtienen para las quimiotecas II y III, respectivamente.

En la Figura 8.16 se muestra la distribución de compuestos para las selecciones SPC (8.16.a) y C.I.D (8.16.b) de la quimioteca I en tres dimensiones que corresponden a los tres primeros PCs, responsables del 69% de la varianza. Puede observarse gráficamente que la selección SPC, que cubre 21 de los 22 *bins* ocupados, contiene grupos de compuestos muy similares entre ellos, mientras que la selección C.I.D, que ocupa menos celdas (17), presenta moléculas mejor espaciadas que uniformemente exploran el espacio.



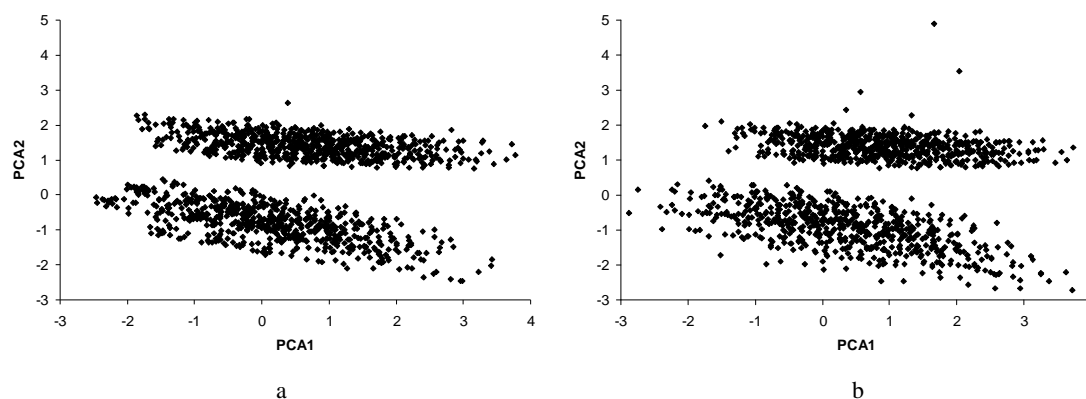
**Figura 8.16.** Selecciones *full array* de la **quimioteca I** en el espacio definido por los tres primeros componentes principales (69% de la varianza). (a) selección 3×2×6 que maximiza la ocupación de celdas (SPC). (b) selección 3×2×6 que minimiza el criterio *cell-integral-diversity*.

Las mismas conclusiones se alcanzan para las dos selecciones de la quimioteca II, mostradas en las Figuras 8.17.a (selección SPC) y 8.17.b (selección C.I.D). Por claridad, se incluye la representación bidimensional de los dos primeros PCs, representando el 63% de la varianza. En este caso, las selecciones SPC y C.I.D ocupan 109 y 82 del total de 147 *bins*. A pesar del menor grado de recubrimiento de la selección C.I.D en términos de celdas ocupadas, esta reducción no se observa visualmente en términos de distancias. Como en el caso de la quimioteca I, la colección C.I.D está más homogéneamente distribuida dentro de los *bins* que la selección SPC, donde los compuestos tienden a agruparse en la región central. La selección SPC rinde un valor de C.I.D de 3.816, mientras que el *score* C.I.D de la quimioteca optimizada con C.I.D es de 3.507 (como referencia, el mínimo valor de C.I.D obtenido para la quimioteca II total es de 3.107).



**Figura 8.17.** Selecciones *full array* de la **quimioteca II** en el espacio definido por los dos primeros componentes principales (63% de la varianza). (a) selección 19×10 que maximiza la ocupación de celdas (SPC). (b) selección 19×10 que minimiza el criterio *cell-integral-diversity*.

Finalmente, en las Figura 8.18.a (selección SPC) y 8.18.b (selección C.I.D) se muestran los gráficos 2-D basados en los dos primeros PCs (52% de la varianza) para las selecciones combinatorias 6×6×6×6 de la quimioteca III. Aquí, la selección optimizada con SPC recubre 423 de las 1209 celdas ocupadas por la quimioteca total, y la selección optimizada con C.I.D, 359 celdas. Los valores C.I.D para cada selección son 2.446 (selección SPC) y 2.325 (selección C.I.D), comparadas con el valor total de 1.648 para C.I.D de la quimioteca III completa. A partir de la inspección visual, la función C.I.D consigue identificar una quimioteca que recubre mejor el espacio que la selección utilizando la métrica SPC.



**Figura 8.18.** Selecciones *full array* de la **quimioteca III** en el espacio definido por los dos primeros componentes principales (52% de la varianza). (a) selección 6×6×6×6 que maximiza la ocupación de celdas (SPC). (b) selección 6×6×6×6 que minimiza el criterio *cell-integral-diversity*.

### 8.3.1.8. Coste computacional

Una de las ventajas de C.I.D sobre D.I es su carácter determinista y menor coste computacional al implicar menos puntos y no necesitar de repeticiones. En este sentido, se registra el esfuerzo computacional para una serie de cálculos en función del tamaño de la selección ( $n$ ) y de las condiciones de la evaluación. En la Tabla 8.18 se muestra el uso de CPU en segundos. Todos los cálculos se ejecutan con un Pentium IV a 2.4 GHz en entorno Linux. El tiempo necesario para realizar la selección no se incluye. Los tiempos mostrados para el método D.I corresponden a una única repetición, por lo que estos valores deberían multiplicarse por el número de repeticiones requerido. El coste computacional asociado con el algoritmo de *Binning* empleado por el método C.I.D es despreciable comparado con los tiempos del método D.I, incluso para la quimioteca III con 531441 compuestos. Para una selección concreta de cardinalidad  $n$ , el uso de CPU de los métodos D.I y C.I.D correlaciona linealmente con el número de puntos. Por ejemplo, para la selección de 1296 compuestos de la quimioteca III evaluada con D.I usando 10000 y 833 puntos, la relación entre el número de puntos (10000/833) y el uso de CPU (1987.04/169.93) es 1.007.

**Tabla 8.18** Tiempo computacional requerido (uso de CPU en segundos) para la evaluación de diversidad por los tres métodos de selecciones de distintas cardinalidades  $n$  de las quimiotecas I, II y III.

	Quimioteca I (N=600)	Quimioteca II (N=35640)		Quimioteca III (N=531441)				
		$n=15$	$n=36$	$n=160$	$n=265$			$n=1296$
SPC	P1 (22 bins)	<0.01	<0.01	P1 (74 bins)	0.06	0.07	P1 (371 bins)	11.36
	P6 (200 bins)	0.01	0.01	P6 (442 bins)	0.13	0.14	P6 (833 bins)	16.47
D.I	22 puntos	0.01	0.01	74 puntos	1.20	1.21	371 puntos	74.21
	200 puntos	0.04	0.05	442 puntos	7.00	7.04	833 puntos	169.93
	1000 puntos	0.23	0.24	10000 puntos	158.62	168.10	10000 puntos	1987.04
C.I.D	P1 (22 puntos)	0.01	0.01	P1 (74 puntos)	1.24	1.27	P1 (371 puntos)	83.14
	P6 (200 puntos)	0.04	0.05	P6 (442 puntos)	7.13	7.21	P6 (833 puntos)	180.90

Por otra parte, en la Tabla 8.19 se muestran los tiempos (uso de CPU) requeridos para optimizar las selecciones *full array* 19×10 y 6×6×6×6 de las quimiotecas II y III. Se observa cómo las optimizaciones con el criterio C.I.D son más lentas que las que utilizan el criterio SPC, ya que las primeras requieren la determinación del vecino más próximo para cada punto situado en el esquema de *binning*.

**Tabla 8.19** Tiempo computacional requerido (uso de CPU en segundos) para optimizar las selecciones de las quimiotecas II y III usando los criterios de SPC y C.I.D con un algoritmo genético.

	Quimioteca II <i>n</i> =190	Quimioteca III <i>n</i> =1296
SPC	7.23 segundos (2400 iteraciones)	6.60 min (2188 iteraciones)
C.I.D	18.12 min (2775 iteraciones)	3.90 horas (2215 iteraciones)

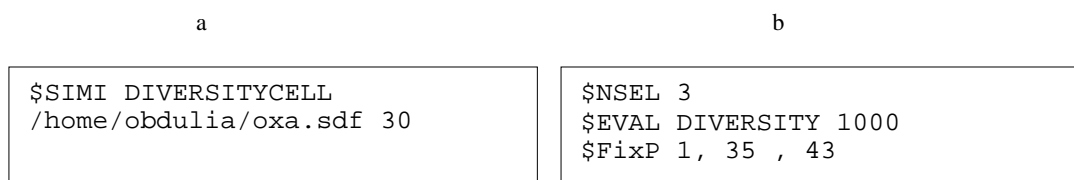
En todos los casos se escoge una población de 30 individuos, por lo que el número de evaluaciones de la función corresponde al producto del número de individuos y el número de iteraciones.

### 8.3.2. Instrucciones para ejecutar en PRALINS evaluaciones de diversidad

Se describen únicamente los comandos para ejecutar C.I.D y D.I, ya que la evaluación *cell-based* se detalla en la referencia [303].

Aunque metodológicamente idénticas, se distingue la posibilidad de comparar dos quimiotecas procedentes de dos ficheros *sd-file* (palabra clave \$SIMI) o de tratarse de una quimioteca contenida en otra, compartiendo el mismo *sd-file* (palabra clave \$EVAL). Tras la palabra clave, se introduce la clave que identifica a uno u otro método: DIVERSITY (D.I) o DIVERSITYCELL (C.I.D). Con la opción \$SIMI, se indica en la línea siguiente el nombre del segundo *sd-file* a compararse, seguido del número de puntos sonda. Con la opción \$EVAL, tras el código DIVERSITY/DIVERSITYCELL se indica el número de puntos. En este caso, los productos que conforman la segunda quimioteca a compararse con la quimioteca total se indican mediante la opción \$FixP, cuyo número debe ser equivalente al número de productos a seleccionarse (\$NSEL). Evidentemente, en las dos opciones las quimiotecas tienen que venir definidas en el mismo espacio químico.

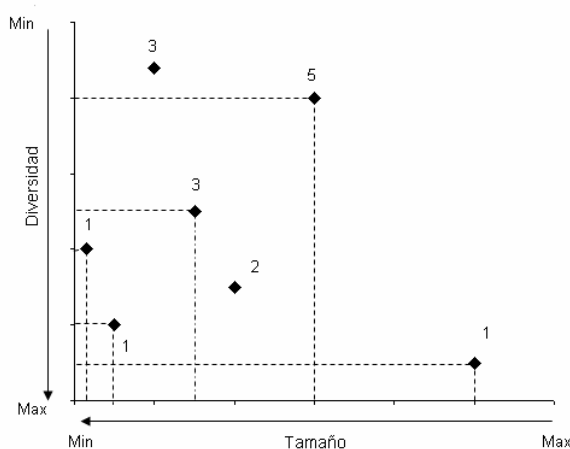
En la Figura 8.19 se muestran dos ejemplos para ejecutar ambas evaluaciones con las dos posibilidades (\$SIMI o \$EVAL).



**Figura 8.19.** Especificaciones para la evaluación de diversidad de (a) quimiotecas procedentes de diferentes *sd-files* con el criterio C.I.D y (b) quimiotecas contenidas en el mismo *sd-file* con el criterio D.I.

## 8.4. Optimización multiobjetivo con algoritmos genéticos (MOGA)

La principal diferencia entre MOGA y GA radica en el modo en que se evalúa la función de *fitness*. Mientras que en GA se analiza un único valor de *fitness*, en MOGA se consideran de manera independiente múltiples objetivos, evaluados bajo el criterio de dominancia de Pareto (apartado 1.9.3). En PRALINS, se considera que un individuo domina a otro si es equivalente o superior en todos los criterios. Cada uno de los individuos recibe un rango de acuerdo al número de veces que es dominado por el resto de individuos de la población. Aquellos no dominados reciben un rango de 1. En la Figura 8.20 se ilustra el método de ranking Pareto en la optimización de dos criterios: minimizar el tamaño de la quimioteca (eje de abscisas) y maximizar diversidad (eje de ordenadas invertido). Para cada uno de los individuos, el rango corresponde al número de individuos comprendidos en el área limitada por el individuo más una unidad.



**Figura 8.20.** Esquema de ranking Pareto para un quimioteca donde se optimizan dos criterios: maximizar diversidad y minimizar tamaño.

La aplicación del criterio de Pareto rinde normalmente un conjunto de soluciones óptimas, conocido como conjunto de óptimos de Pareto (*Pareto Optimal set*), constituido por soluciones no dominantes. La representación de los valores de las funciones objetivo para este conjunto de soluciones se denomina frente de Pareto (*Pareto front*). La complejidad de la jerarquización de Pareto es  $O(kM^2)$  donde  $k$  es el número de funciones objetivo y  $M$  es el tamaño de la población.

Los posibles objetivos  $f(S)$  para un conjunto  $S$  de  $n$  compuestos que contempla PRALINS son:

- Criterio de diversidad. Por el momento implementado exclusivamente para los métodos basados en distancias.
- Propiedades contenidas en el *sd-file*, donde se puede maximizar/minimizar:
  - El valor mínimo o máximo de la propiedad en el conjunto seleccionado.
  - La suma de valores de la propiedad en el conjunto seleccionado. Por ejemplo, puede desearse minimizar el coste total de compra de reactivos de síntesis de la quimioteca.
  - El promedio de valores de la propiedad  $x_i$  en el conjunto seleccionado. Por ejemplo, la propiedad puede definirse como la actividad frente a una diana terapéutica calculada con un modelo QSAR o *docking*, dirigiéndose el diseño hacia su maximización (ecuación [8.11])

$$f(S) = \frac{1}{n} \sum_{i=1}^n x_i \quad [8.11]$$

- El confinamiento de la propiedad  $x_i$  entre unos valores mínimo ( $x_{\min}$ ) y máximo ( $x_{\max}$ ) (ecuación [8.12]). Si se desea que las propiedades alcancen un valor objetivo particular  $x^*$ , la ecuación anterior se re-escibe como la ecuación [8.13]. En ambos casos, se obtienen quimiotecas restringidas minimizando este criterio.

$$f(S) = \frac{1}{n} \sum_{i=1}^n \max(x_{\min} - x_i, x_i - x_{\max}, 0) \quad [8.12]$$

$$f(S) = \frac{1}{n} \sum_{i=1}^n \text{abs}(x_i - x^*) \quad [8.13]$$

- Número total de reactivos a utilizarse en un diseño *full array*. MOGA se programa por el momento para el diseño *full array*. La codificación de los cromosomas se realiza en bits, permitiéndose en este caso que el tamaño y la configuración del *full array* varíen durante la optimización. Para ello, el usuario indica el límite inferior y superior de reactivos a seleccionar para cada uno de los puntos de variación. Tanto en el momento de generar la población aleatoria inicial como tras los eventos de *crossover* y mutación, únicamente se permiten individuos que satisfagan los límites impuestos (ambos inclusive). El resto son obligados a mutar hasta satisfacer la restricción. Para ejecutar MOGA con una configuración concreta, basta con igualar ambos límites. En el momento en que estos límites difieren, PRALINS analiza cualquier configuración contenida en ellos, independientemente de si la opción de minimizar el número total de reactivos está activada o no. Es decir, en el caso de optimizar únicamente diversidad, PRALINS explorará las configuraciones contenidas en los límites, pero únicamente devolverá aquella que maximice este único criterio. Si además se activa el criterio de minimizar el número total de reactivos, devolverá todas aquellas configuraciones que maximicen diversidad y minimicen el número total de reactivos necesarios.

Como métodos de selección, ya que la función de *fitness* se calcula en rangos, se considera únicamente el operador *Roulette Wheel selection* con una ordenación de los valores de la función objetivo proporcional al rango. Como método de *crossover* se aplica el *single-point crossover*.

Las técnicas de optimización multiobjetivo basadas en el criterio de Pareto requieren de un método que permita mantener la diversidad en la población. Para reducir la tendencia a la deriva génica inducida por el operador de selección se aplican técnicas de *niching* (apartado 1.9.2.5). En este caso, se aplica la técnica *fitness sharing*<sup>292</sup>, de manera análoga a como está implementada en el programa MOSELECT.<sup>159,295</sup> En cada nueva generación, se evalúa para los individuos no dominados (rango 1) la proximidad de otras soluciones no dominantes. La primera solución no dominante constituye el centro de un *niche*. Si la siguiente solución no dominante presenta el valor de las funciones objetivo dentro de un radio del valor de las funciones objetivo de soluciones que componen un *niche*, se penaliza dicha solución incrementándose su rango en una unidad. De lo contrario, constituye un nuevo *niche*. El valor de este radio o *niche radius* se calcula según un porcentaje para cada uno de los valores de las funciones objetivo que presenta la solución del centro del *niche*. Dicho porcentaje es un parámetro definido por el usuario, que por defecto adopta un valor del 1%.

El reemplazo es condicional, de manera que tras aplicarse el criterio de Pareto y *niching* al conjunto de soluciones (población parental y filial), la nueva generación se compone de los mejores individuos.

Al no disponerse de un único valor a ser evaluado como en el caso de GA, no es trivial establecer un criterio de convergencia. Así, se fija como criterio de terminación el número máximo de iteraciones permitido. En el fichero de salida de PRALINS se recogen todas las soluciones del conjunto óptimo de Pareto, estén contenidas o no en un mismo *niche*. El proceso de la optimización se recoge en el fichero *pointsmulti.txt*, en el que se imprimen los valores de las funciones objetivo para cada uno de los individuos de la población filial en cada iteración.

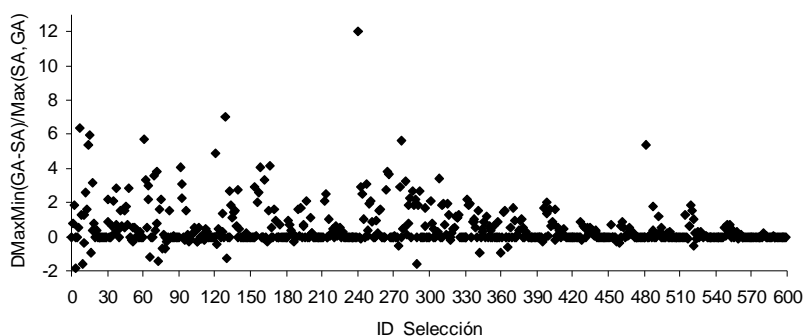
Como en las selecciones *full array* bajo un único criterio de diversidad, se permite la fijación de una serie de reactivos para cada punto de diversidad. El concepto de grupo de reactivos comunes en distintos puntos de diversidad no se introduce.

#### 8.4.1. Optimización multiobjetivo de diversidad y número de reactivos

Se investiga el comportamiento de MOGA en la identificación de selecciones que maximicen el criterio de diversidad MaxMin promediado (ecuación [1.110]) y minimicen el número de reactivos a seleccionarse.

El primer caso de estudio es la quimioteca I (apartado 8.2.2.1) de 600 oxazoles ( $5 \times 4 \times 30$ ). Como referencia para MOGA, se realizan 600 selecciones independientes con GA optimizando únicamente el criterio MaxMin promediado, cada una de las cuales corresponde a una de las 600 posibles configuraciones de selección *full array*. Los parámetros de GA utilizados son: tamaño de población de 30, 2000 iteraciones máximo (100 *idle*), una tasa de *crossover* de 0.8 y de mutación de 0.6. Como método de selección se aplica el *Roulette Wheel selection* con escalado de la función de *fitness* y como método de *crossover* el *single-point crossover*. El reemplazo es condicional, sin repetición de genomas.

Se repiten también estas 600 selecciones con SA para comparar la eficiencia y eficacia de GA frente a SA. De las 600 optimizaciones, en 277 selecciones GA rinde selecciones con valores de MaxMin promediado superiores a SA, en 278 casos ambos algoritmos son equivalentes y en 45 casos SA rinde valores superiores a GA. En la Figura 8.21 se representa la relación entre la diferencia de valores MaxMin promediado obtenidos con GA y SA y el valor máximo obtenido con cualquiera de estos dos métodos para cada una de las selecciones. El eje de abcisas corresponde a la identidad de selección, ordenadas éstas siguiendo la secuencia:  $1 \times 1 \times 1$ ,  $1 \times 1 \times 2, \dots, 1 \times 1 \times 30, 1 \times 2 \times 1, \dots, 5 \times 4 \times 29, 5 \times 4 \times 30$ .

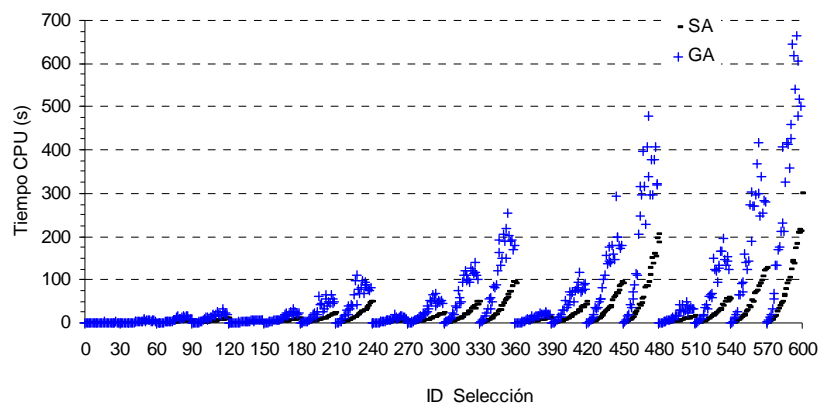


**Figura 8.21.** Porcentaje entre la diferencia de valores MaxMin promediado obtenidas con GA y SA relativo al valor máximo obtenido con cualquiera de los dos métodos. El eje de abcisas corresponde a la identidad de selección, estando éstas ordenadas siguiendo la secuencia  $1 \times 1 \times 1, 1 \times 1 \times 2, \dots, 1 \times 1 \times 30, 1 \times 2 \times 1, \dots, 5 \times 4 \times 29, 5 \times 4 \times 30$ .

Respecto a la eficiencia, el tiempo total de uso de CPU para realizar las 600 selecciones es de 11939 segundos y 37691 segundos para SA y GA, respectivamente. Todos los cálculos se

ejecutan en un Pentium IV a 2.4 GHz en entorno Linux. En la Figura 8.22 se muestra el tiempo de CPU necesario para cada una de las selecciones, ordenadas como se ha expuesto.

Se observa cómo a medida que se requieren configuraciones más próximas al número total de reactivos disponibles para cada punto de variación, el tiempo se incrementa para ambos algoritmos y especialmente para GA, debido a la mutación interna que obliga a alcanzar un tamaño de selección fijado.



**Figura 8.22.** Tiempo de CPU (s) requerido para cada una de las 600 selecciones con GA y SA.

Así, para las condiciones de convergencia ajustadas en cada algoritmo, GA es en general más efectivo que SA, aunque menos eficiente, especialmente para las situaciones indicadas, donde será preferible utilizar SA.

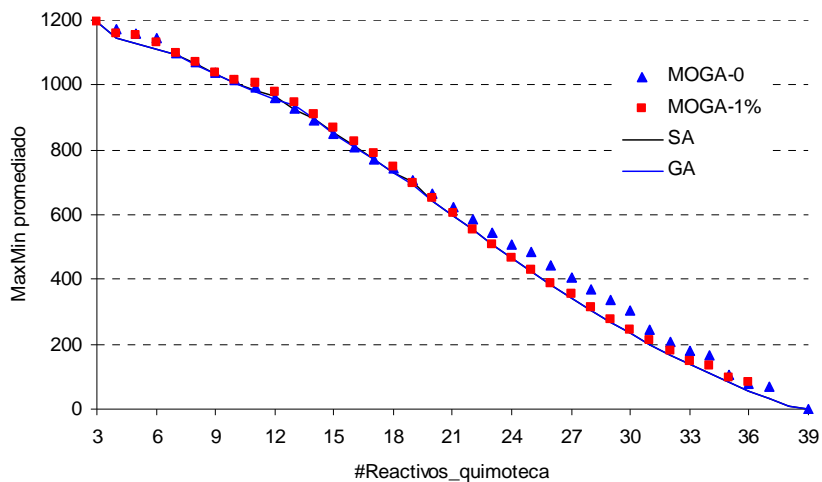
Sobre estos 600 resultados se aplica el criterio de Pareto (minimizar el número de reactivos y maximizar diversidad) para identificar el conjunto de óptimos de Pareto, separadamente para GA y SA. En ambos casos, éste consta de 35 selecciones posibles, cuyo frente de Pareto se muestra en la Figura 8.23. En el eje de abscisas se muestra el criterio de número de reactivos a minimizar y en el eje de ordenadas el criterio MaxMin promediado. Para facilitar su interpretación, se representan los valores referidos al valor máximo de MaxMin (correspondiente a la selección de toda la quimioteca  $5 \times 4 \times 30$ ), con lo que el sentido de los dos criterios se optimiza en el origen de coordenadas.

Por otro lado, se ejecuta un único cálculo MOGA que optimice ambos criterios, con unos parámetros de tamaño de población de 100, 300 iteraciones, una tasa de *crossover* de 0.8, una tasa de mutación de 0.6 y sin *niching* (MOGA-0). Dicho cálculo se repite con *niche radius* del 1% (MOGA-1%). MOGA sin *niching* identifica los 35 óptimos de Pareto, considerando como referencia el *set* de 35 encontrado con los cálculos individuales de GA + aplicación del criterio de Pareto. Estos 35 óptimos se muestran en la Figura 8.23 (*triángulos azules*). Así la optimización multiobjetivo funciona satisfactoriamente desde el punto de vista de identificación de los posibles óptimos. MOGA con un *niche radius* del 1% (Figura 8.23, *cuadrados rojos*) identifica 34 de los 35 óptimos, perdiendo un óptimo correspondiente a la configuración  $5 \times 4 \times 28$  (corresponde a la de 37 reactivos en el eje de abscisas que falta en la Figura 8.23). La pérdida es probablemente debida a la aplicación del propio *niching*, ya que excluye durante la optimización soluciones próximas a las presentes (por ejemplo, la configuración  $5 \times 4 \times 30$  que requiere 39 reactivos). Es común que la aplicación de técnicas de *niching* disminuya el número de soluciones óptimas de Pareto, disminución proporcional al *niche radius* aplicado.

Sin embargo, prescindiendo de la pérdida de una solución óptima, se observa cómo la introducción de *niching* permite que el valor de diversidad de las soluciones se aproxime al obtenido con GA y SA, comparativamente frente a la situación de no introducir *niching*.



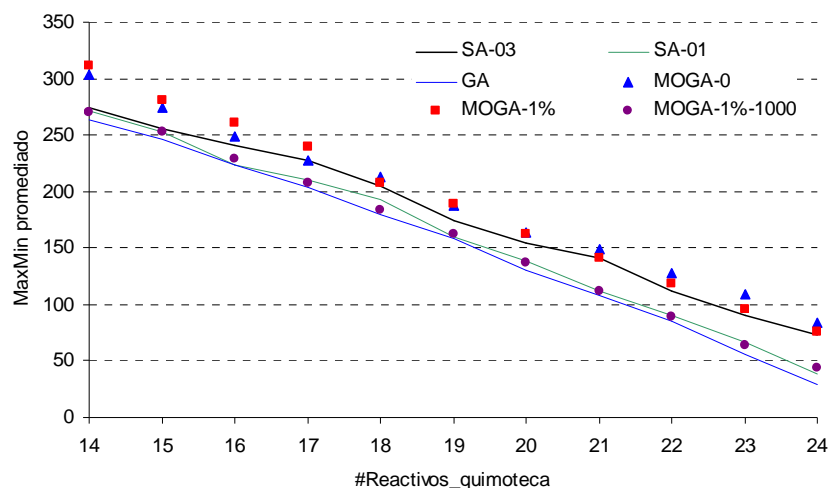
El tiempo de CPU requerido para ambas optimizaciones multiobjetivo es de 155 segundos (MOGA-0) y 776 segundos (MOGA-1%). Considerando la eficiencia computacional del proceso, el tiempo total se reduce 77 veces (MOGA sin *niching*) y 15 veces (MOGA con *niching*) respecto al cálculo independiente con SA (eficiencia aún mayor si se compara con GA).



**Figura 8.23.** Conjunto de óptimos de Pareto encontrados con 600 ejecuciones independientes de GA (línea azul), 600 ejecuciones independientes de SA (línea negra), una única ejecución MOGA sin *niching* (triángulos azules) y una única ejecución con MOGA y *niche radius* del 1% (cuadrados rojos). Los dos objetivos son minimizar el número total de reactivos (eje de abscisas) y maximizar la diversidad (eje de ordenadas). El eje de ordenadas se encuentra invertido de manera que los valores se refieren al valor máximo obtenido (quimioteca total). Así, los dos criterios se optimizan en el origen de coordenadas.

Como segundo caso de estudio se escoge la quimioteca II, BIB\_Amino de 35640 (165×216) compuestos, descrita según el apartado 8.2.2.1, optimizándose los dos mismos objetivos que en el caso anterior. En este caso se generan de manera independiente con GA y SA las 36 quimiotecas con configuraciones comprendidas en los márgenes 7×7 y 12×12 (ambos inclusive). Para GA, los parámetros son similares a los de la quimioteca anterior, excepto que el número máximo de iteraciones se fija en 5000 (200 *idle*). Para SA, se prueban dos criterios de parada, el tradicional, con fracción de aceptación inferior a 0.3 (SA-03), y con fracción de aceptación inferior a 0.1 (SA-01).

La optimización MOGA se realiza con los mismos parámetros anteriores y en condiciones i) sin *niching* (MOGA-0) ii) con radio de *niching* de 1% (MOGA-1%) y con iii) con radio de *niching* de 1% y extendiendo el número de iteraciones de 300 hasta 1000 (MOGA-1%-1000). En la Figura 8.24 se muestran el frente de Pareto obtenido en las seis situaciones, con un total de 11 soluciones óptimas identificadas en todos los casos. Dado el elevado coste computacional asociado al cálculo MaxMin promediado para el total de la quimioteca, se toma 450 como valor de referencia máximo para invertir el eje de ordenadas.



**Figura 8.24.** Conjunto de óptimos de Pareto encontrados con 36 ejecuciones independientes de GA (línea azul), 36 ejecuciones independientes de SA-03 (línea negra) y SA-01 (línea verde), una única ejecución MOGA sin *niching* (triángulos azules), una única ejecución con MOGA y *niche radius* del 1% con 300 iteraciones (cuadrados rojos) y una única ejecución con MOGA y *niche radius* del 1% con 1000 iteraciones (puntos violetas). Los dos objetivos son minimizar el número total de reactivos (eje de abcisas) y maximizar la diversidad (eje de ordenadas). El eje de ordenadas se encuentra invertido de manera que los valores se refieren a un valor máximo (establecido de 450). Así, los dos criterios se optimizan en el origen de coordenadas.

Se observa de nuevo cómo la eficacia de GA frente a SA es superior, aún imponiéndose un criterio de parada para SA más estricto (fracción de aceptación de 0.1). En este caso, es necesario alargar MOGA hasta 1000 iteraciones para obtener soluciones lo más próximas al cálculo independiente con GA/SA-01. En la Tabla 8.20 se listan los tiempos totales de uso de CPU para realizar las selecciones con cada una de las seis situaciones. Se observa la mayor eficiencia de SA-03 frente a GA. De nuevo, considerando la eficiencia y la eficacia, MOGA con 1000 iteraciones es muy rentable.

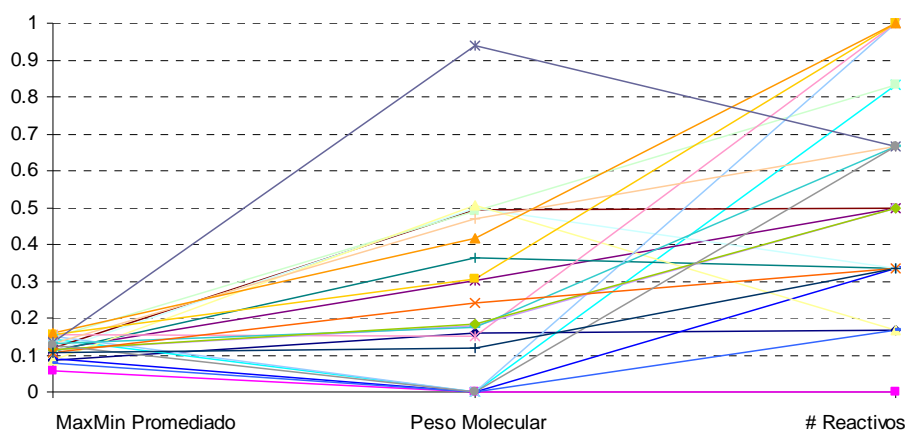
**Tabla 8.20.** Tiempo computacional requerido (uso de CPU en segundos) para realizar el total de 36 selecciones (GA, SA-01, SA-03) y una única ejecución MOGA en la quimioteca II.

Método	Tiempo CPU (s)
SA-01	339
SA-03	1001
GA	4706
MOGA-0 300 iteraciones	136
MOGA-1% 300 iteraciones	153
MOGA-1% 1000 iteraciones	531

#### 8.4.2. Optimización multiobjetivo de varias propiedades

En este apartado se ilustra el comportamiento de MOGA en la optimización de más de dos objetivos: i) maximizar diversidad con el criterio MaxMin promediado, ii) minimizar número de reactivos a seleccionarse y iii) el peso molecular de los compuestos de la quimioteca debe comprender valores entre 300 y 500 g/mol. El cálculo se realiza sobre la quimioteca de oxazoles, explorando únicamente aquellas quimiotecas con configuraciones comprendidas entre  $2 \times 2 \times 2$  y  $4 \times 4 \times 4$  (ambos límites inclusive). El número de iteraciones se fija en 1000, con un *niche radius* del 1% y manteniéndose el resto de parámetros como en el apartado 8.4.1.

Se obtienen 23 soluciones no dominantes, mostradas en la Figura 8.25. Las funciones objetivo se muestran en el eje de abscisas y cada solución se muestra como una línea continua. Los valores de las funciones objetivo se escalan entre 0 y 1. Para el criterio MaxMin promediado, se identifica como máximo de 1 el valor de la quimioteca total, por lo que es normal que las selecciones aquí representadas se encuentren debajo de este valor sin alcanzar el máximo de 1. El confinamiento de peso molecular se escala entre 0 (todos los compuestos debidamente confinados) y el valor máximo de la quimioteca total (que presenta compuestos fuera de los extremos), por lo que en este caso las quimiotecas óptimas tenderán a 0. El número de reactivos se escala entre el mínimo posible en este caso: 6 (2+2+2) y el valor máximo de 12 (4+4+4), por lo que también las quimiotecas óptimas tienden a 0.



**Figura 8.25.** Quimiotecas encontradas al optimizar simultáneamente diversidad, número de reactivos y confinamiento del peso molecular entre 300 y 500.

En la Tabla 8.21 se incluyen una serie de selecciones con valores de la función objetivo en los extremos (mostrados en negrita), indicando que MOGA ha convergido ya que explora los límites. El cálculo tarda 19 segundos de tiempo de CPU.

**Tabla 8.21.** Selecciones escogidas del conjunto de óptimos de Pareto de la Figura 8.23.

Configuración	MaxMin promediado	Confinamiento MW	# Reactivos
2×2×2	<b>68.3</b>	<b>0</b>	<b>6</b>
4×4×4	<b>188.1</b>	10.8	<b>12</b>
3×3×4	160.4	<b>24.4</b>	10
4×4×4	180.9	<b>0</b>	<b>12</b>

### 8.4.3. Instrucciones para ejecutar MOGA en PRALINS

La palabra clave es \$MOGA seguida del número de propiedades del *sd-file* a optimizarse y seguida de la opción YES/NON (por defecto) en caso de que se quiera minimizar o no el número total de reactivos de la quimioteca. El criterio de diversidad se encuentra siempre activado. En la línea siguiente se especifican los parámetros del algoritmo genético:

```
GA_POPULATION GA_ITERATION GA_CROSSOVER GA_MUTATION GA_NICHING
```

Donde GA\_NICHING corresponde al porcentaje de radio de *niching* y el resto de parámetros coinciden con los expuestos en el apartado 8.1.1.

En las líneas siguientes, se listan los nombres de las propiedades a optimizarse seguidas del carácter ':', y los cuatro códigos numéricos que definen el tipo de optimización para esta propiedad (Tabla 8.22): i) si es maximización/minimización, ii) el tipo de definición (máximo, mínimo, suma, promedio, confinamiento a un valor o en un margen), iii) el límite inferior permitido para esta propiedad y iv) el límite superior permitido. Aunque no se requiera confinamiento, los valores de los límites deben indicarse con ceros. Si se carece de alguna especificación, PRALINS devuelve error.

**Tabla 8.22.** Códigos para la definición de las funciones objetivo.

Códigos para determinar el sentido de optimización de una propiedad	
1	Minimización
2	Maximización

Códigos para la definición del criterio de una propiedad	
1	Mínimo
2	Máximo
3	Suma
4	Promedio (ecuación [8.11])
5	Confinamiento a un valor (ecuación [8.13])
6	Valor confinado en un límite inferior y superior (ecuación [8.12])

Finalmente, se deben fijar los límites inferior y superior del número de reactivos a seleccionar para cada punto de variación bajo las palabras clave \$NSELLOW y \$NSELUP, respectivamente. Estas opciones invalidan la opción \$NSEL usada en optimizaciones con un único criterio. El criterio de diversidad se establece según la palabra clave \$SUBDISI, como en la versión original de PRALINS.

En la Figura 8.26.a se muestra el fichero de entrada para la optimización del criterio MaxMin, de la maximización del valor promedio de la actividad calculada (ACT\_CAL), el ajuste a un valor de 4.5 para el número de aceptores (H\_bond acceptor) y el confinamiento del peso molecular (MW) entre 400 y 700. En la Figura 8.26.b se muestra el correspondiente fichero de entrada para maximizar el criterio MaxMin promediado y minimizar el número total de reactivos, explorando todas las configuraciones comprendidas entre 1×1×1 y 5×4×30.

a	b
<pre> \$SUBDISI MXMIN \$MOGA 3 NON 100 5000 0.8 0.5 0.01 ACT_CAL: 2 4 0 0 H_bond acceptor: 1 5 4.5 MW: 1 6 400 700 \$NSELLOW 4: 2 1 2 \$NSELUP 24: 3 2 4                     </pre>	<pre> \$SUBDISI MXMAV \$MOGA 0 YES 100 300 0.8 0.6 0.01  \$NSELLOW 1: 1 1 1 \$NSELUP 600: 5 4 30                     </pre>

**Figura 8.26.** Formato del fichero de texto para especificar en PRALINS selecciones multiobjetivo.

## 8.5. Otras implementaciones

### 8.5.1. Ampliación de los métodos de *clustering*

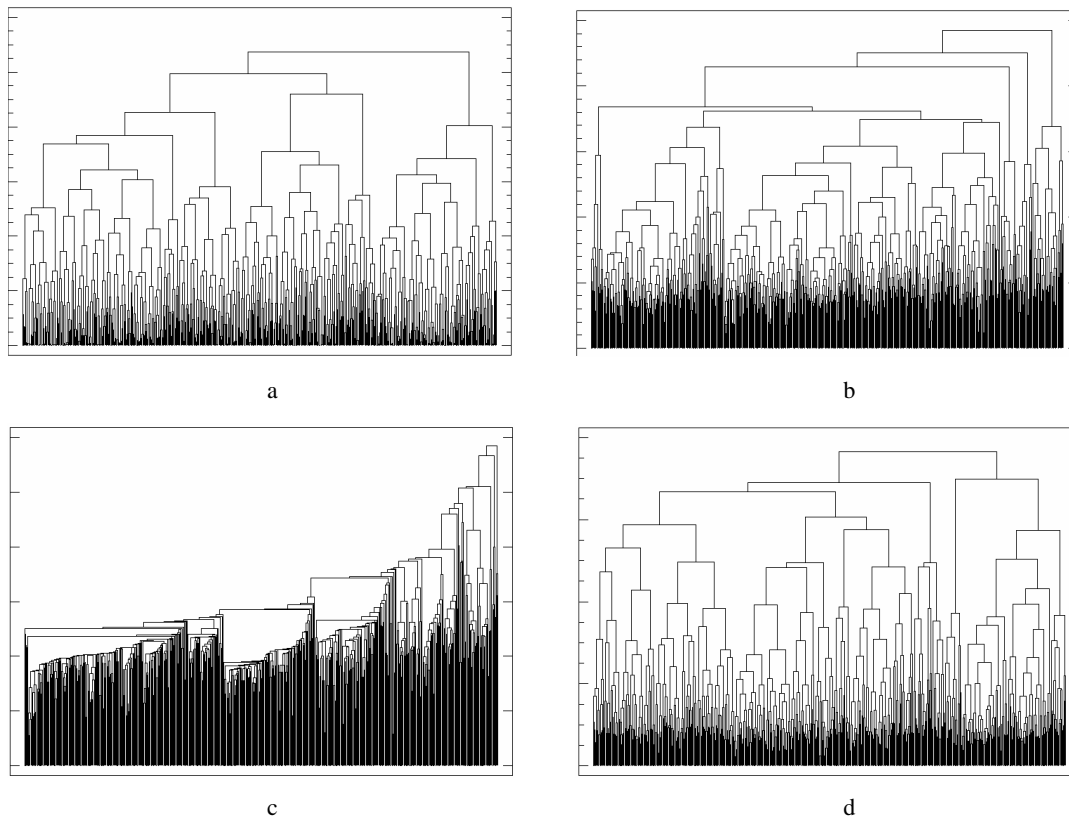
Dentro de los métodos de *clustering* jerárquico aglomerativo se añaden las variantes Group Average y Ward, simplemente introduciendo las constantes necesarias en la fórmula de Lance-Williams (ecuación [1.113]). Las selecciones de diversidad (*cherry picking* y *full array*) basadas en *clustering jerárquico* se ejecutan a través de la palabra HRC seguida del código que indica la variante correspondiente. En la Tabla 8.23 se muestran las palabras clave necesarias en PRALINS para cada tipo de método. Se añade a la descripción el nombre correspondiente usado en programas comerciales como Cerius2 o el programa STATGRAPHICS<sup>464</sup>, ya que existen pequeñas diferencias en la nomenclatura o en la función implementada (caso del método de Ward en Cerius2).

**Tabla 8.23.** Códigos necesarios en PRALINS para cada variante de HRC. Comparación con la nomenclatura utilizada en STATGRAPHICS y Cerius2.

	$\alpha$	$\beta$	$\delta$	$\gamma$	PRALINS	Cerius2	STAT
Complete linkage (Furthest Neighbour)	0.5	0.5	0	0.5	complete	Complete linkage	Furthest Neighbour
Median linkage	0.5	0.5	-0.25	0	median	NO	Median
Single linkage (Nearest Neighbour)	0.5	0.5	0	-0.5	single	Single linkage	Nearest neighbor
Centroid	$\frac{n_i}{n_i + n_j}$	$\frac{n_j}{n_i + n_j}$	$\frac{-n_i \cdot n_j}{(n_i + n_j)^2}$	0	average	NO	Centroid
Average linkage (weighted) or Group Average	$\frac{n_i}{n_i + n_j}$	$\frac{n_j}{n_i + n_j}$	0	0	groupaverage	Average linkage	Group Average
Ward	$\frac{n_i + n_k}{n_i + n_j + n_k}$	$\frac{n_j + n_k}{n_i + n_j + n_k}$	$-\frac{n_k}{n_i + n_j + n_k}$	0	ward	Ward's*	Ward's

\*en Cerius2, el método de Ward utiliza el cuadrado de la distancia Euclídea, por lo que también se habilita esta métrica en PRALINS bajo la palabra clave EUCD2. STAT: STATGRAPHICS.

El método de Ward ofrece la ventaja de que permite una distribución más homogénea de los datos en los *clusters*. En la Figura 8.27 se muestran los dendrogramas correspondientes a las dos variantes introducidas para el *clustering* de los 600 oxazoles en el espacio definido por los 7 componentes principales de la quimioteca I.



**Figura 8.27.** Dendrogramas para los métodos de *clustering* (a) Ward, (b) Group Average, (c) Single y (d) Complete.

Se añaden también otros dos métodos de *clustering* jerárquicos aglomerativos basados en la densidad (*density linkage*): KNN y Uniformity Kernel. Ambos proceden básicamente generando una nueva medida de disimilitud basada en estimaciones de densidad y adyacencias, en función de la cual se caracterizan los métodos. A partir de esta matriz de disimilitudes se realiza un *cluster* siguiendo la fórmula del *single linkage*.<sup>465</sup>

- KNN (*k*-th-Nearest neighbor Method). La matriz de distancias se calcula siguiendo el siguiente procedimiento:

-Para un punto  $x$  calcula la distancia  $r_k(x)$  desde este punto a su  $k$ -th vecino más próximo, siendo  $k$  un parámetro introducido por el usuario. Se considera una esfera centrada en  $x$  con radio  $r_k(x)$ , de manera que la densidad  $f(x)$  estimada en  $x$  corresponde a la proporción de observaciones dentro de la esfera dividido por el volumen de la esfera (ecuación [8.14]):

$$f(x) = \frac{k}{n \cdot V_k(x)} \quad [8.14]$$

-Para un par de puntos  $x_i$  y  $x_j$  se calcula la nueva interdistancia según la ecuación [8.15]:

$$d^*(x_i, x_j) = \left\{ \begin{array}{ll} \frac{1}{2} \left( \frac{1}{f(x_i)} + \frac{1}{f(x_j)} \right) = \frac{n}{2 \cdot k} [V_k(x_i) + V_k(x_j)] & \text{si } d(x_i, x_j) \leq \max(r_k(x_i), r_k(x_j)) \\ \infty & \text{si no son vecinos} \end{array} \right\} \quad [8.15]$$

- *Uniformity Kernel*. Procede de manera análoga al anterior, aunque en este caso el usuario determina un radio  $r$  de *kernel* idéntico para todos los compuestos. Para un punto  $x$ , la densidad  $f(x)$  corresponde a la proporción de observaciones dentro de la esfera dividido por el volumen de la esfera de radio  $r$ . Para un par de puntos  $x_i$  y  $x_j$  se calcula la nueva interdistancia según la ecuación [8.16].

$$d^*(x_i, x_j) = \left. \begin{cases} \frac{1}{2} \left( \frac{1}{f(x_i)} + \frac{1}{f(x_j)} \right) = \frac{n}{2 \cdot k} [V_k(x_i) + V_k(x_j)] & \text{si } d(x_i, x_j) \leq r \\ \infty & \text{si no son vecinos} \end{cases} \right\} \quad [8.16]$$

Para ejecutar estos métodos en PRALINS debe utilizarse la palabra clave que indica el formato de la selección de diversidad (\$DISI para *cherry picking* o \$SUBDISI para quimiotecas combinatorias), seguido del nombre del método (KNN o UniKernel) y el parámetro que requiere cada uno de ellos ( $k$  o  $r$ , respectivamente). El resto de opciones son las estándar para los correspondientes métodos de *clustering* (referencia [303]). En la Figura 8.28 se muestran ejemplos para cada uno de ellos.

a	b
<pre>\$DISI KNN 4 RND</pre>	<pre>\$SUBDISI UniKernel 0.9 3 pop 0.9</pre>

**Figura 8.28.** Comandos en modo texto para enviar cálculos (a) KNN y (b) *Uniformity Kernel*.

### 8.5.2. Lectura de *fingerprints* procedentes de MOE

Se incorpora la lectura de *fingerprints* procedentes del MOE en PRALINS:

- *Fingerprints* farmacofóricos. Se habilita únicamente la métrica Tanimoto usada en MOE para estos descriptores.
- La versión empaquetada en bits del *fingerprint* MACC, calculados con la función `ph4_MACCS_Fingerprint`. Se habilitan todas las métricas.

Tras el cálculo de los *fingerprints* en MOE, debe exportarse un fichero en formato ASCII que contenga únicamente el campo de estos descriptores, sin incluir títulos y sin comillas. Cada una de las líneas corresponde a un compuesto, cuyos valores del *fingerprint* deben ir separados por comas.

Para su lectura en PRALINS, debe introducirse la palabra clave \$BINVAR seguida de la abreviatura correspondiente a la métrica a emplear. En la línea siguiente se especifica el tipo de *fingerprint*: `FGP_4pharMOE` (*fingerprint* farmacofórico) o `FGP_MACCSmoe` (MACCS) y a continuación el nombre del fichero de *fingerprints* con el *path* completo.

### 8.5.3. Métricas en los métodos de clasificación

Junto a la optimización de la fracción de espacio y población, se incorpora la posibilidad de optimizar los criterios *cell-based Chi2* (ecuación [1.105]) y *cell-based Entropy* (ecuación [1.106]) en los dos métodos de clasificación (*binning* y *clustering*). En *binning*, se habilita para todos los métodos de optimización (Local Search, MTC, SA y GA), en *clustering*, únicamente para GA.

Además, para los métodos de *binning* se incorpora el *cell-integral-diversity criterion* presentado en el apartado 8.3.

Los correspondientes códigos para estas métricas se listan en la Tabla 8.24.

**Tabla 8.24.** Códigos para la definición de las métricas en métodos de clasificación.

---

ent	<i>cell-based Entropy</i>
chi	<i>cell-based Chi2</i>
cid	<i>cell-integral-diversity criterion</i>

---



## Capítulo 9.

### Estudio de los co-receptores CXCR4 y CCR5.

En este capítulo se presenta un estudio de la modelización de los co-receptores CXCR4 y CCR5 y el modo de unión de sus inhibidores. Dicho trabajo se encuentra enmarcado en la línea de investigación del grupo en el campo de inhibidores de fusión del virus de la inmunodeficiencia humana (HIV). En este proyecto, participan el departamento de Química Orgánica y Bioquímica del Instituto Químico de Sarriá y el laboratorio de retrovirología de la Fundación IrsiCaixa en el Hospital Germans Trias i Pujol.

En los últimos años, la quimioterapia del HIV se ha dirigido hacia nuevas dianas farmacológicas, más allá de las tradicionales como la transcriptasa reversa y la proteasa.<sup>466</sup> Entre ellas, destacan los fármacos que actúan inhibiendo la entrada del virus a la célula, que presentan la evidente ventaja de impedir la infección, así como una acción potencial contra la resistencia del HIV, causar menos efectos secundarios y permitir su administración en un régimen de dosis más simple.<sup>467</sup>

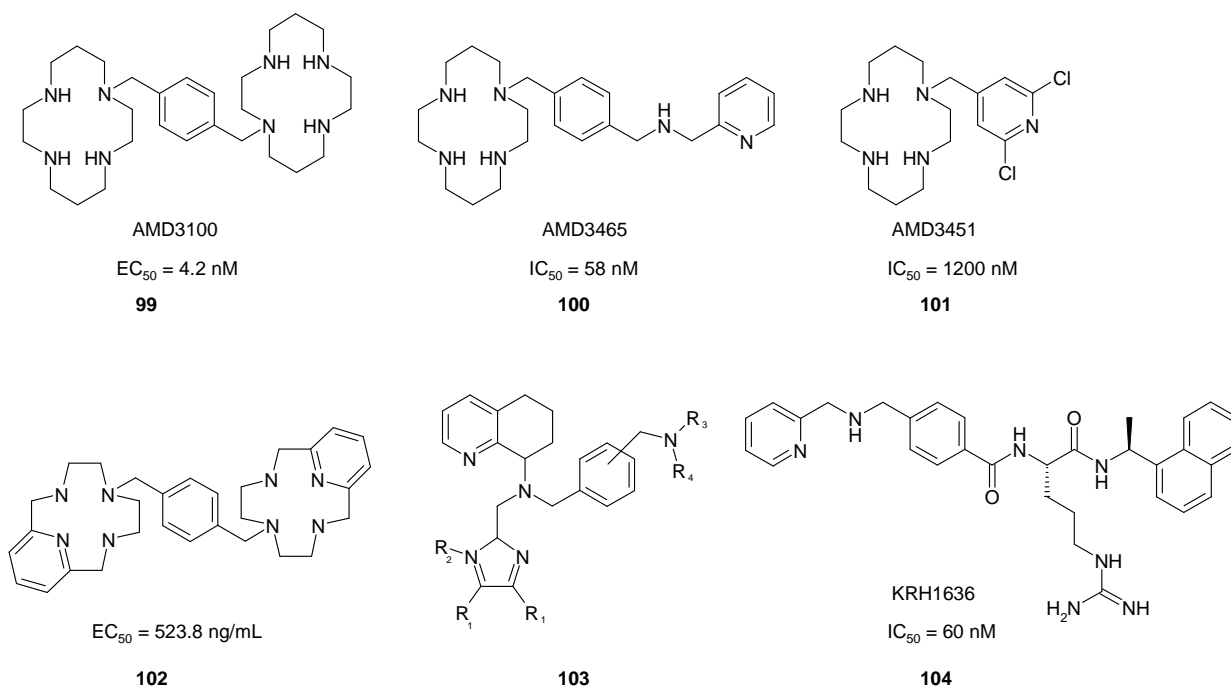
El HIV se fusiona con las células CD4 (aquellas que presentan esta glicoproteína, principalmente) a través de la interacción del *loop* V3 de la glicoproteína gp120 de su envoltura con CD4, unión que provoca un cambio conformacional en gp120, exponiendo un sitio de unión hacia los co-receptores CXCR4<sup>468</sup>/CCR5<sup>469</sup> (mayoritariamente estos). El descubrimiento de la participación de estos receptores en el proceso de fusión permitió explicar el tropismo de las cepas de HIV, y desde entonces se han convertido en una diana terapéutica prominente.<sup>470,471</sup> El tropismo, preferencia por la infección de ciertas células, resulta interesante desde el punto de vista terapéutico. En las etapas iniciales de establecimiento de la enfermedad, el virus ataca principalmente a monocitos y macrófagos (M-trópico), siendo CCR5 el co-receptor fundamentalmente implicado (cepas R5). En la fase sintomática en que la enfermedad progresa, las cepas infectan preferentemente a los linfocitos T, utilizando el co-receptor CXCR4 (cepas X4 o T-trópicas). También pueden presentarse cepas duales (R5X4). Así, en principio, resulta más interesante bloquear CCR5, por su papel fundamental en la transmisión. Además, a diferencia de CCR5, la ausencia o bloqueo de CXCR4 no es tan bien tolerada debido a su función en la hematopoyesis.<sup>472</sup>

CXCR4 y CCR5 son receptores de quimiocinas pertenecientes a la familia de los receptores acoplados a la proteína G (GPCRs). La característica funcional común a las GPCRs es su mecanismo de transmisión de señales a través de la disociación de las unidades de la proteína G. Estructuralmente, se caracterizan por siete hélices  $\alpha$  que atraviesan la membrana celular, conectadas alternativamente por *loops* intra y extracelulares. El extremo N-terminal tiene localización extracelular y el C-terminal, citoplasmática. Concretamente, los receptores de quimiocinas pertenecen a la familia A (*Rhodopsin-like*) de GPCRs, que presenta un elevado grado de conservación de ciertos residuos en los dominios transmembrana (TM).<sup>229</sup>

## 9.1. Inhibidores antagonistas del co-receptor CXCR4

Dentro de los inhibidores más comunes de CXCR4 destacan <sup>470-471</sup>:

- Moléculas orgánicas pequeñas.
  - ◆ Derivados **biciclamos**. Desarrollados por AnorMED, que se encuentran en ensayos clínicos de fase II/III. <sup>473-475</sup> Estos compuestos fueron inicialmente descritos en el año 1992 <sup>476</sup> como inhibidores selectivos de la replicación del HIV tipo I y II. Posteriormente, se describe su papel como antagonistas de CXCR4. <sup>477</sup> El cabeza de serie es el AMD3100 (**99**) (Figura 9.1). Esta serie se caracteriza por dos anillos macrocíclicos (1,4,8,11-tetraazaciclodecano) conectados por un fragmento aromático o alifático. El AMD3100 es un antagonista puro y específico de la CXCR4, de elevada potencia. Sin embargo, su baja biodisponibilidad causada por su carga total positiva a la par que ciertas anomalías cardíacas, han motivado el desarrollo de una nueva serie de inhibidores administrables por vía oral.
  - ◆ Derivados **monociclamos** como AMD3465 <sup>478</sup> (**100**) o AMD3451 <sup>479</sup> (**101**) (Figura 9.1). Todavía no son suficientemente biodisponibles, pero también son potentes antagonistas de CXCR4 e incluso CCR5 (AMD3451).
  - ◆ Derivados de **azamacrociclos** <sup>480</sup> como los derivados de 1,10-fenantrolina, 2,2'-bipiridilo o derivados de bis-piridil macrociclos (**102**) (Figura 9.1) cuyo tamaño de anillo varía entre 12 y 16 miembros por anillo.
  - ◆ Derivados de **tetrahydroquinolinamina** <sup>481-485</sup> desarrollados por AnorMED, en los cuales se reemplaza la estructura de ciclamos por N-(1H-benzimidazol-2-ilmetil)-5,6,7,8-tetrahydro-8-quinolinamina (**103**) (Figura 9.1).
  - ◆ Derivados de **KRH1636** <sup>486-488</sup> (**104**), antagonista con una potencia similar a AMD3100. (Figura 9.1).



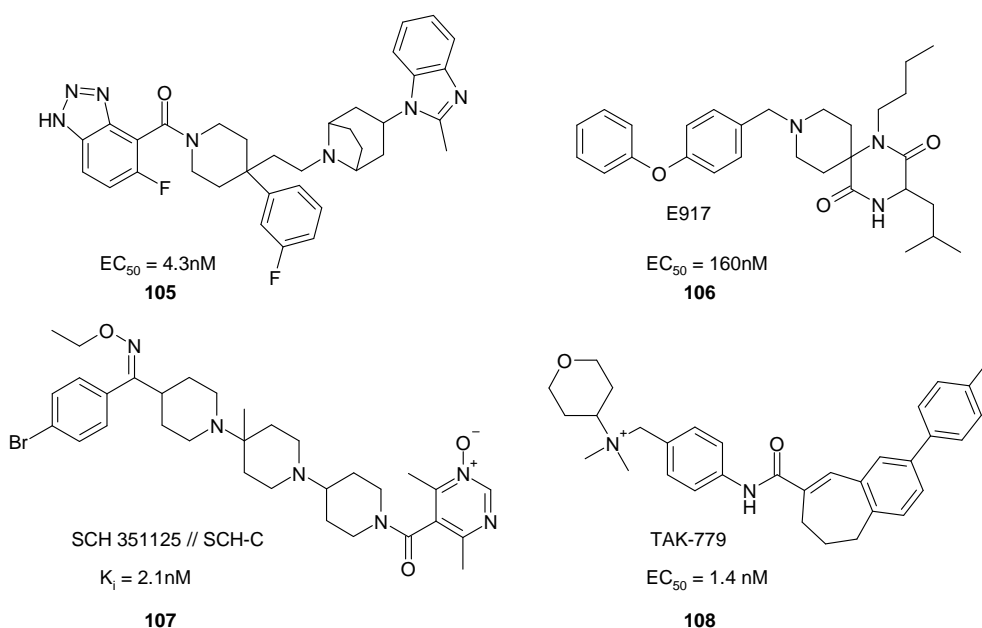
**Figura 9.1.** Estructuras de moléculas orgánicas pequeñas antagonistas de CXCR4.

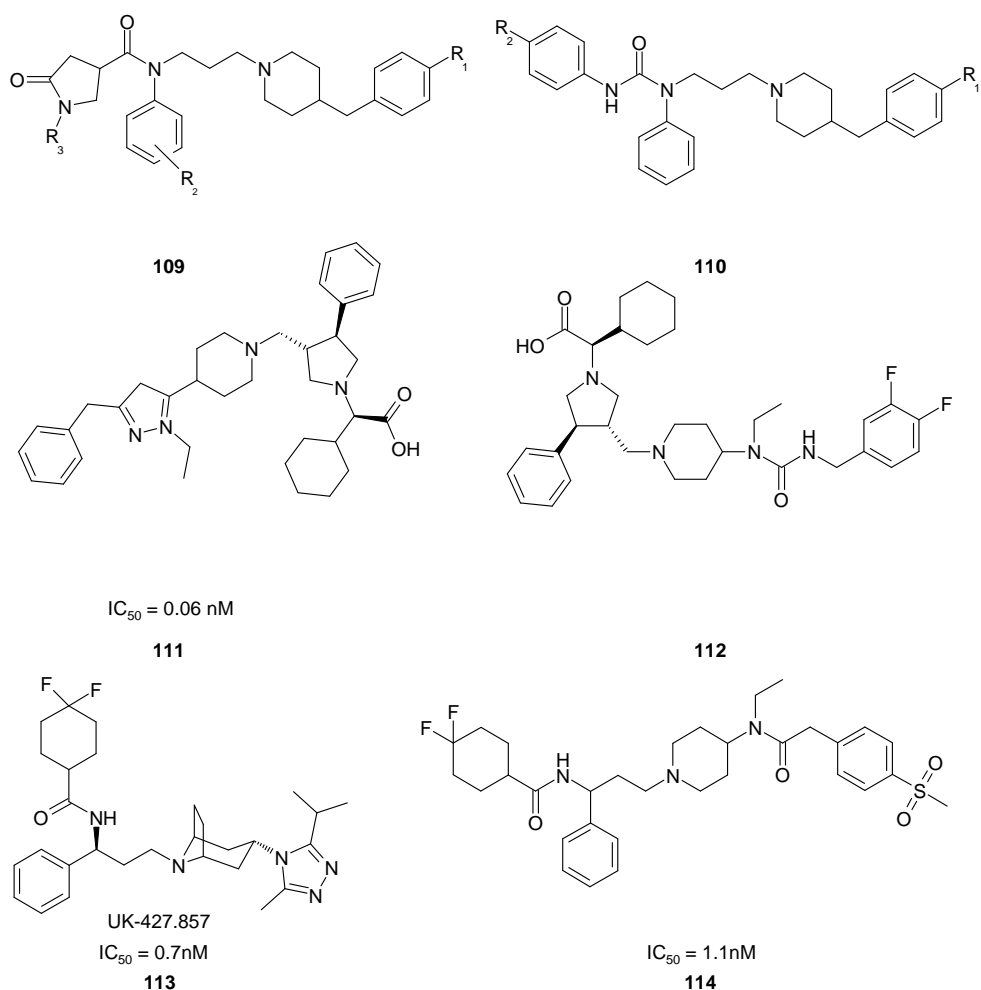
- Antagonistas peptídicos. Como el *lead* T22<sup>489</sup> y sus derivados de menor peso molecular T134<sup>490</sup> y 14-mer T140<sup>491</sup> desarrollados a partir de un péptido antimicrobial natural (polifemusinas). Destaca también ALX40-4C.<sup>492</sup>
- Agentes basados en quimiocinas. Tanto quimiocinas naturales como modificaciones, por ejemplo de SDF-1 $\alpha$ , ligando natural de CXCR4.

## 9.2. Inhibidores antagonistas del co-receptor CCR5

Dentro de los inhibidores más comunes de CCR5 destacan:

- Moléculas orgánicas pequeñas (Figura 9.2):
  - ◆ Derivados de **fenilciclohexilamina**<sup>493-498</sup> (**105**)
  - ◆ Derivados de **dicetopiperazina**<sup>499-503</sup> (**106**)
  - ◆ Derivados de **SCH-C**<sup>504-506</sup> (**107**)
  - ◆ Derivados de **TAK 779**<sup>507-508</sup> (**108**)
  - ◆ Derivados de **5-oxopirrolidina-3-carboxamida**<sup>509</sup> (**109**)
  - ◆ Derivados de **Urea**<sup>510</sup> (**110**)
  - ◆ Derivados de **1,3,4-pirrolidina-piperidina trisustituida**<sup>511</sup> (**111**)
  - ◆ Derivados de **anillos pentacíclicos 1,3,5 trisustituidos**<sup>512</sup> (**112**)
  - ◆ Derivados de **1-fenil-1,3-propanodiamina**. Cuyo representante, el UK-427.857 se encuentra en fase clínica II.<sup>513-515</sup> (**113**)
  - ◆ Derivados de **4-amino-piperidina o tropano**<sup>516-518</sup> (**114**)
  - ◆ Derivados de **4-piperidina**<sup>518-519</sup>
  - ◆ Derivados de **anilida N-óxido de piridina**<sup>520</sup>
- Anticuerpos monoclonales. Como PRO 140, en fase clínica I.
- Quimiocinas modificadas. Derivados de RANTES, ligando natural de CCR5.
- Antagonistas peptídicos. Como el MKK-1.





**Figura 9.2.** Estructuras de moléculas orgánicas pequeñas antagonistas de CCR5.

### 9.3. Datos bioquímicos de la interacción de AMD3100 con CXCR4

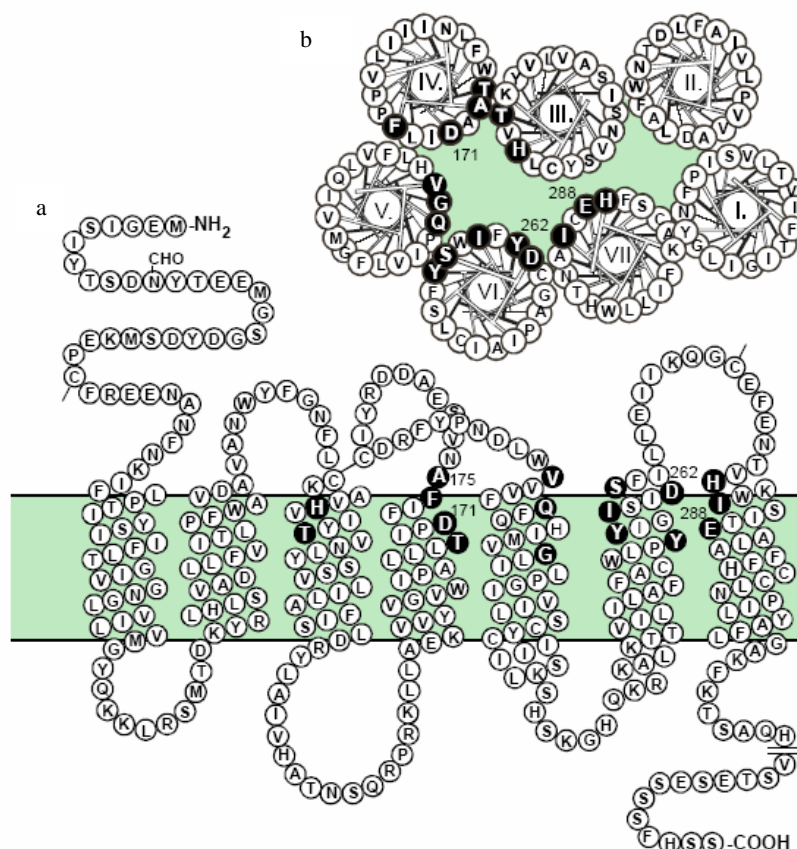
De todos los antagonistas de CXCR4 presentados en el apartado 9.1, se estudia en profundidad la interacción de la familia de compuestos biciclamo con dicho receptor, concretamente para el AMD3100 (**99**), ya que es para el que existen más datos experimentales de mutagénesis dirigida y receptores quiméricos.

Dada la carga electrostática del AMD3100 a pH fisiológico (+2) y su capacidad quelante de cationes como Zn<sup>+2</sup> o Ca<sup>+2</sup>, los experimentos de mutagénesis dirigida se han centrado en los residuos ácidos (Asp y Glu) y en las histidinas situadas en los *loops* extracelulares y segmentos TM. También se ha analizado la influencia de posibles impedimentos estéricos, en los que Ala, Gly o Ile han sido mutados por Phe o Trp. En la Figura 9.3 se muestra un diagrama bidimensional y helicoidal de la estructura de CXCR4, indicándose aquellos residuos identificados como importantes en la unión del AMD3100 con CXCR4.<sup>521-525</sup> Destacan:

- Asp<sup>171</sup>. Situado en la TM4.
- Asp<sup>262</sup>. Situado en la TM6.
- Glu<sup>288</sup>. Situado relativamente próximo a la superficie celular, participa en la unión del SDF-1 $\alpha$  y se hipotetiza que puede interactuar con el *loop* V3 de la glicoproteína gp120 del HIV.

- Ala<sup>175</sup>. Situado al inicio del segundo *loop* extracelular ECL2, frente a la cavidad formada por los residuos anteriores. La mutación de Ala<sup>175</sup> por Phe genera impedimento estérico.
- Tyr<sup>255</sup>. Situado en la TM6.
- His<sup>281</sup>. Situada en el extremo de TM7, su mutación por Ala potencia el efecto antiviral y antagonista de AMD3100.

La sustitución de los residuos Asp<sup>171</sup>, Asp<sup>262</sup> y Glu<sup>288</sup> por Ala o Asn reduce tanto la actividad antiviral del AMD3100 como su efecto antagonista frente a SDF-1 $\alpha$ .<sup>522-524</sup> Además, Asp<sup>171</sup> y Asp<sup>262</sup> son también esenciales para la función de CXCR4 como co-receptor del HIV.<sup>523</sup>

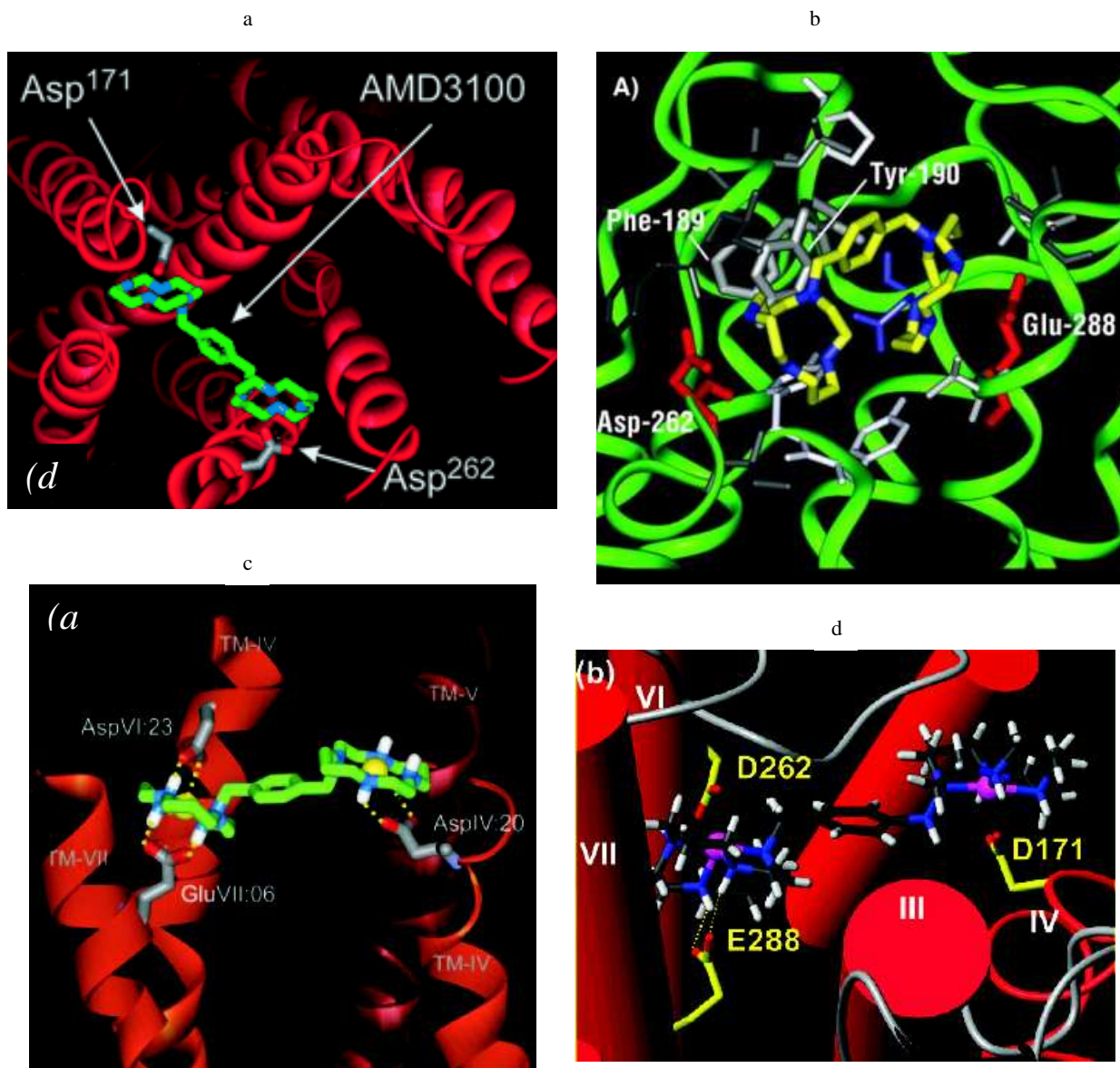


**Figura 9.3.** (a) Diagrama bidimensional y (b) Diagrama helicoidal de CXCR4. Los residuos en negro indican residuos mutados. Se destaca la posición del Asp<sup>171</sup>, Asp<sup>262</sup> y Glu<sup>288</sup>. Extraído de [521].

Con estos resultados, se plantea un modo de interacción electrostático entre el AMD3100/AMD3100(Zn<sub>2</sub>) y los residuos ácidos, de manera que cada anillo de ciclamo interaccione a través de sus nitrógenos con el grupo carboxilato correspondiente. Basándose en estudios computacionales, diferentes autores han propuesto diversas alternativas de unión.

- Modelo propuesto por Schwartz<sup>524</sup> en el que cada ciclamo interacciona con Asp<sup>171</sup> y Asp<sup>262</sup>, respectivamente (Figura 9.4.a).
- Modelo propuesto por Trent<sup>526</sup> (Figura 9.4.b) en el que cada ciclamo interacciona con Asp<sup>262</sup> y Glu<sup>288</sup>.

- Modelo propuesto por Schwartz<sup>525</sup> (Figura 9.4.c) y Sadler<sup>527</sup> (Figura 9.4.d). Ambos autores coinciden en un modelo tipo “sándwich” en el que uno de los anillos de ciclamato interactúa con el Asp<sup>171</sup> y el otro conjuntamente con Asp<sup>262</sup> y Glu<sup>288</sup>. Este modelo se propone para el compuesto AMD3100(Zn<sub>2</sub>). En el modelo propuesto por Sadler, la configuración del ciclamato que interactúa con el par Asp<sup>262</sup> y Glu<sup>288</sup> es *cis-V*, estableciéndose un enlace Zn-COO<sup>-</sup> (2.27 Å) con Asp<sup>262</sup> y un enlace por puente de hidrógeno (COO<sup>-</sup>...NH, 2.01 Å) con el Glu<sup>288</sup>. El otro ciclamato se encuentra en configuración *trans-I*, con una distancia entre el Zn y el grupo carboxilato de 2.88 Å.



**Figura 9.4.** (a) Modelo de interacción propuesto por Schwartz, extraído de [524]. (b) Modelo de interacción propuesto por Trent, extraído de [526]. (c) Modelo tipo “sándwich” propuesto por Schwartz para la unión AMD3100(Zn<sub>2</sub>) extraído de [525]. (d) Modelo tipo “sándwich” propuesto por Sadler para AMD3100(Zn<sub>2</sub>), extraído de [527].

#### 9.4. Datos bioquímicos de la interacción de TAK-779 con CCR5

TAK-779 es uno de los inhibidores de CCR5 del que más información experimental del sitio de unión, obtenida mediante mutagénesis dirigida, se dispone. TAK-779 (**108**) (Figura 9.2) es la primera molécula no peptídica identificada que inhibe la función del receptor CCR5.<sup>528</sup> Presenta la desventaja de requerir administración intravenosa.

Los estudios indican que TAK-779 (**108**), así como SCH-C (**107**) (Figura 9.2), se unen principalmente en una cavidad común situada en el dominio transmembrana de CCR5, a la par que otros elementos de compuestos específicos establecen contactos adicionales.<sup>529-531</sup>

Las estructuras de TAK-779 y SCH-C presentan grupos aromáticos y regiones hidrofílicas, como un átomo de nitrógeno positivamente cargado. Mientras que TAK-779 contiene un ión amonio cuaternario, SCH-C presenta un nitrógeno de piperidina, protonado a pH fisiológico. En principio, se podría suponer que las partes hidrofóbicas de TAK-779 y SCH-C interaccionan con el dominio TM de CCR5, mientras que el átomo de nitrógeno cargado positivamente, lo haría con la región extracelular. Sin embargo, esta hipótesis no es soportada por los resultados experimentales de mutagénesis dirigida, que no aportan evidencia de interacción con residuos ácidos de la región extracelular.<sup>531</sup>

Concretamente para TAK-779, se han identificado los siguientes residuos, localizados en TM1, TM2, TM3 y TM7<sup>529</sup> (Figura 9.5).

- Glu<sup>283</sup>. El grupo carboxilato actúa de contraión para el nitrógeno con carga positiva. Además, no existen en CCR5 contraiones próximos a este aminoácido, lo que facilitaría la interacción iónica con los inhibidores. Su sustitución por alanina tiene un efecto intermedio en la actividad antiviral de TAK-779.
- Residuos con un fuerte efecto inhibidor de la actividad antiviral del TAK-779: Leu<sup>33</sup>, Tyr<sup>37</sup>, Trp<sup>86</sup>, Tyr<sup>108</sup>, Thr<sup>123</sup>.
- Residuos con un efecto inhibidor intermedio de la actividad antiviral del TAK-779: Arg<sup>31</sup>, Thr<sup>82</sup>, Ile<sup>198</sup>.
- Residuos con un efecto inhibidor mínimo: Phe<sup>79</sup>, Leu<sup>104</sup>.

Se han realizado estudios de *docking*<sup>532</sup> que confirman como residuos más importantes en la interacción al Glu<sup>283</sup>, Tyr<sup>37</sup>, Trp<sup>86</sup> y Tyr<sup>108</sup>. La participación de otros residuos encontrados en *docking* como Thr<sup>195</sup> o Leu<sup>255</sup> no se confirma mediante mutagénesis dirigida.

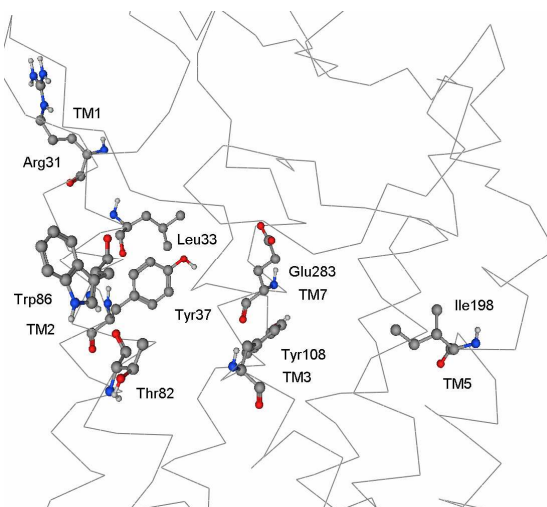


Figura 9.5. Cavidad de unión de TAK-779 a CCR5.

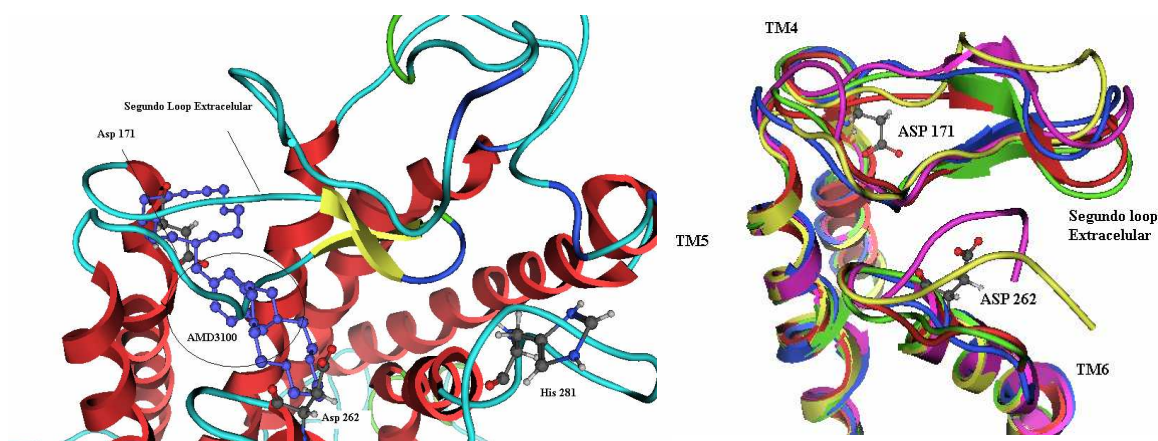


## 9.5. Modelos de CXCR4 y CCR5

En trabajos anteriores realizados en el grupo<sup>303</sup> se modelaron las estructuras de los co-receptores CXCR4 y CCR5 por homología a partir de la estructura de la rodopsina bovina, con el programa MODELLER. Como se ha comentado en el apartado 1.5.1, dada la falta de patrones estructurales para GPCRs, el uso del mapa de rayos-X de la rodopsina bovina se ha impuesto en el campo de la modelización de GPCRs. Concretamente, se utilizó la entrada PDB 1HZX<sup>533</sup>, con resolución de 2.80 Å.

La modelización de los segmentos transmembrana se realizó completamente por homología, sin *gaps* y forzando el alineamiento de los motivos (*patterns*) y residuos más conservados. La modelización de los *loops* se realizó *ab initio* con el programa MODELLER, exceptuando los extremos N-terminal, C-terminal y el *loop* extracelular ECL2. Este *loop*, que conecta la TM4 y TM5, presenta un motivo de estructura secundaria  $\beta$ -*hairpin* que se considera conservado en todas las GPCRs de la familia A.<sup>534</sup> Dado que en la modelización *ab initio* del *loop* ECL2 se perdía esta estructura secundaria, se optó por mantener el segmento modelado por homología a partir del alineamiento con la rodopsina bovina.

CXCR4 presenta dos puentes disulfuro entre Cys<sup>109</sup>-Cys<sup>186</sup> y entre Cys<sup>28</sup>-Cys<sup>274</sup>. Particularmente, resulta interesante el primer puente, conservado en la rodopsina bovina, que une a la Cys<sup>186</sup> (situada en el *loop* ECL2) con la Cys<sup>109</sup> (situada en el extremo de TM3 y ECL1). Tal y como se observa en la Figura 9.6, este *loop* cierra la cavidad de unión del AMD3100 descrita. En el trabajo anteriormente realizado, se observó la dificultad de posicionar manualmente el AMD3100 en la cavidad de unión en presencia del *loop* ECL2 modelado con el programa MODELLER, encontrándose un elevado número de *bad contacts* (Figura 9.6.a). Con el fin de obtener una mayor apertura de este *loop*, se realizaron una serie de dinámicas moleculares a diferentes temperaturas, ninguna de las cuales consiguió un gran cambio conformacional debido a la presencia del puente disulfuro (Figura 9.6.b).



**Figura 9.6.** (a) *Docking* manual del AMD3100 en CXCR4. (b) Comparación de la estructura del segundo *loop* extracelular de CXCR4, ECL2, en las sucesivas dinámicas moleculares realizadas.<sup>303</sup>

Uno de los primeros objetivos de este estudio es el de incrementar la apertura del *loop* ECL2, a fin de reproducir la entrada del AMD3100. Para ello, se utiliza el programa CONGEN (*CON*formation *GEN*erator)<sup>535</sup>, que muestrea de manera exhaustiva los diferentes grados de libertad del *loop*: torsiones de la cadena principal y cadenas laterales y sus extremos. Concretamente, se aplica el procedimiento de búsqueda directa (*direct search*) que es el que permite modelar *loops* de mayor longitud.



A partir del modelo obtenido en trabajos anteriores (en adelante denominado MODEL\_MODELLER), se modelan consecutivamente los *loops*, comenzando por los de mayor longitud y de más externos a internos respecto a su posición en la proteína.<sup>536</sup> De todos los conformeros obtenidos para un *loop*, se escoge el que presente una mayor RMSD respecto al modelo inicial, siempre y cuando su energía sea como máximo 10kcal/mol superior a la del conformero de mínima energía, ya que posteriormente se minimiza la estructura global obtenida.

Los residuos de aminoácidos que comprenden los *loops* se listan en la Tabla 9.1, considerando la predicción de Gerlach<sup>524</sup> de segmentos TM basados en hélices  $\alpha$  utilizada en la modelización. Se muestra también la RMSD de superposición de todos los átomos para cada segmento modelado con CONGEN respecto a la estructura inicial.

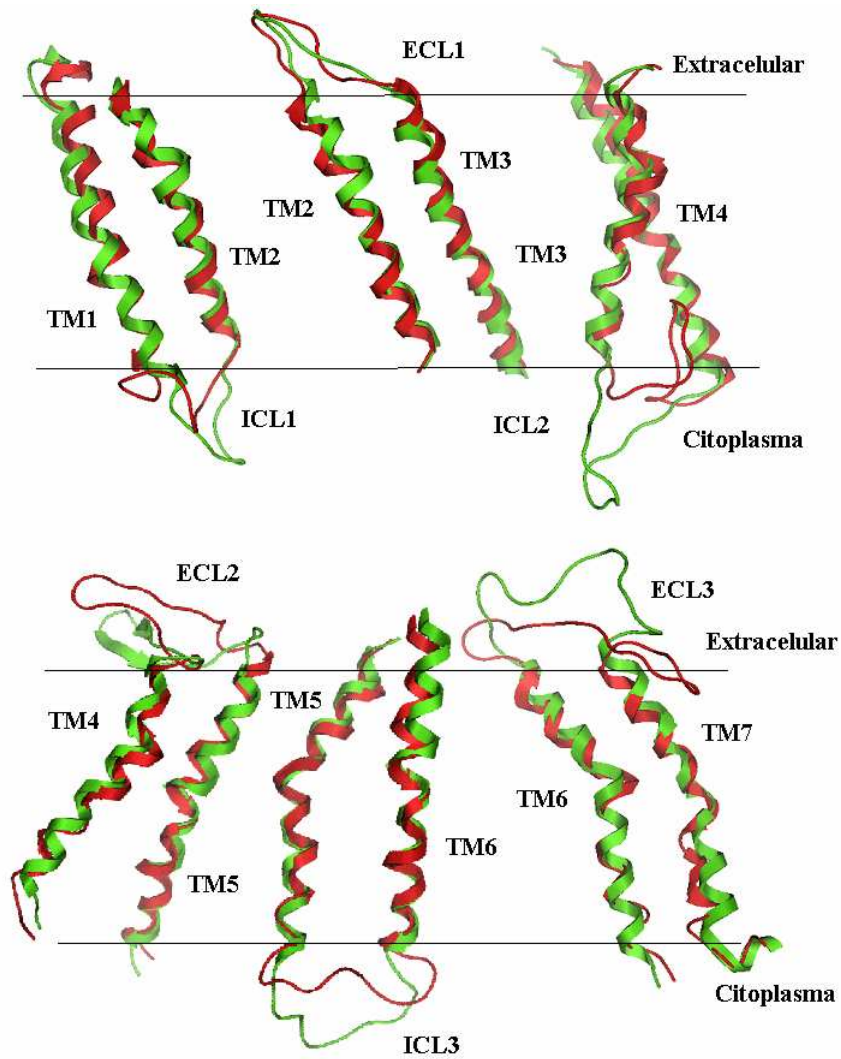
**Tabla 9.1.** Residuos que comprenden los 6 *loops* extra e intracelulares en CXCR4 y su RMSD respecto al modelo inicial.

	ICL1	ECL1	ICL2	ECL2	ICL3	ECL3
Residuos	65-74	98-108	136-151	175-195	224-235	264-281
RMSD (Å)	5.97	5.17	7.54	8.01	5.42	7.34

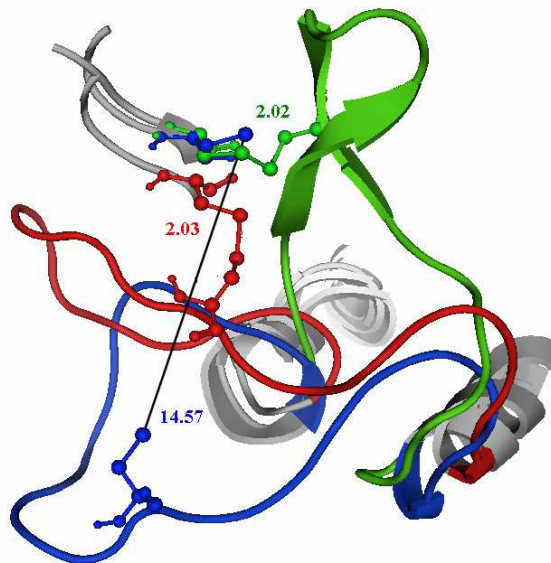
En la Figura 9.7 se muestra la superposición de cada uno de los *loops* modelados *ab initio* con CONGEN, en adelante denominado MODEL\_CONGEN (*rojo*), respecto a la conformación inicial MODEL\_MODELLER (*verde*).

CONGEN no mantiene el puente disulfuro Cys<sup>109</sup>-Cys<sup>186</sup>, ni siquiera añadiendo la restricción respectiva a este tipo de enlaces, restricción que únicamente consigue acercar ambos residuos respecto a la situación de no incluirla. Por ello, se decide editar manualmente el correspondiente fichero *pdb* e imponer la conexión de ambos residuos, minimizándose el *loop* ECL2 con el *force field* Amber94 en el programa MOE. En la Figura 9.8 se ilustra la superposición del *loop* ECL2 de MODEL\_MODELLER (*verde*), con el modelo obtenido con CONGEN sin mantener el puente disulfuro (*azul*) y el modelo final de CONGEN imponiendo dicho puente (*rojo*). Se observa cómo se disminuye la distancia desde 14.57 Å hasta 2.03 Å al imponer la conexión de dichas cisteínas y cómo el *loop* ECL2 se sitúa más alejado de la cavidad de unión del AMD3100 si se compara con MODEL\_MODELLER. Así, pese a que se pierde el motivo  $\beta$ -*hairpin*, dado que se consigue esta apertura, se opta por incluir este modelo en estudios de *docking* del AMD3100.

La generación de los *loops* con CONGEN se realiza por el momento para CXCR4, ya que es la primera diana para la que se habían realizado anteriormente estudios de *docking* y dinámica molecular.



**Figura 9.7.** Superposición de los *loops* de MODEL\_CONGEN (rojo) frente a MODEL\_MODELLER (verde).



**Figura 9.8.** Detalle de la superposición del *loop* ECL2 en MODEL\_MODELLER (verde), el modelo obtenido con MODEL\_CONGEN (azul) en el que no se mantiene el puente disulfuro y el modelo obtenido con MODEL\_CONGEN (rojo) al que se impone la conexión Cys<sup>109</sup>-Cys<sup>186</sup> seguido de una minimización.

Previamente a la realización de estudios de *docking* para estos modelos, se valida la estructura MODEL\_CONGEN de CXCR4 y MODEL\_MODELLER de CCR5 mediante el programa PROCHECK.<sup>537</sup> (Tabla 9.2). Ambos modelos satisfacen los requisitos estereoquímicos.

**Tabla 9.2.** Resumen de la validación de los modelos de CXCR4 y CCR5 con PROCHECK.

	MODEL_CONGEN CXCR4	MODEL_MODELLER CCR5
Porcentaje de residuos con ángulo diedros de la cadena principal en las regiones del diagrama de Ramachandran	64.2% regiones favorables 29.3% regiones permitidas 5.0% poco permitidas 1.6% no permitidas	71.2% regiones favorables 26.3% regiones permitidas 1.5% poco permitidas 0.9% no permitidas
Número de residuos Gly / Pro en regiones no favorables de los diagramas de Ramachandran de estos aminoácidos	2 de 29 residuos	2 de 27 residuos
Número de residuos en conformaciones no favorables según mapas $\chi_1$ - $\chi_2$ (Arg, Asn, Asp, Cys, Gln, Glu, His, Ile y Leu)	7 de 218 residuos	7 de 226 residuos
Factor G global indicador de la estereoquímica global del modelo. Idealmente el valor debe estar comprendido entre -0.5 y -1.	-0.6	-0.5
Número de impedimentos estéricos	0	0
Porcentaje de residuos con distancias de enlace en la cadena principal en los límites permitidos.	100%	100%
Porcentaje de residuos con ángulos de enlace en la cadena principal en los límites permitidos	74%	77.4%
Porcentaje de residuos con grupos planares en la cadena lateral en los límites permitidos	88.2%	77.5%

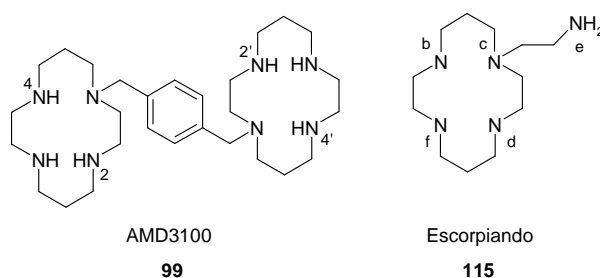
## 9.6. Estudio del sitio y modo de unión del AMD3100 en CXCR4

En este apartado se realiza la predicción del modo de unión del AMD3100 en CXCR4 mediante *docking*, tanto orientado en la cavidad de unión (apartado 9.3) como por *docking* ciego.

Se considera un pH de trabajo de 7, manteniéndose los residuos ácidos desprotonados y los básicos protonados. Las conclusiones de la mutagénesis dirigida de His<sup>281</sup> sugieren que este residuo presenta repulsión electrostática con el anillo de ciclamó (carga +2), por lo que se decide protonar dicha histidina. El resto de histidinas, a falta de datos experimentales, se mantienen neutras.

El anillo de ciclamó del AMD3100 presenta unas constantes de protonación de los cuatro nitrógenos de 10.92, 10.51, 2.65 y 1.62. Para establecer la secuencia de protonación, se adopta como referencia la secuencia de protonación del escorpiando<sup>538</sup> (**115**), en la que en primer lugar se protona el N<sub>d</sub>, seguido del N<sub>b</sub>, N<sub>e</sub>, N<sub>f</sub>, y finalmente N<sub>c</sub> (Figura 9.9). Con ello, las estructuras

de biciclamos con las que se trabaja están protonadas en los nitrógenos 2, 4 y 2', 4' del AMD3100 (Figura 9.9).



**Figura 9.9.** Estructura del escorpiando y secuencia de protonación del AMD3100.

Los diferentes estudios de *docking* realizados con AUTODOCK son:

- Predicción del sitio de unión del AMD3100 en CXCR4 mediante *docking* ciego. El objetivo es determinar si la modificación del *loop* ECL2 realizada permite o no simular la entrada del AMD3100 en la cavidad de unión, por lo que el ligando se posiciona inicialmente fuera de dicha cavidad de unión.
- Predicción del modo de unión mediante *docking*, en el que el AMD3100 se sitúa inicialmente en la cavidad de unión descrita en el apartado 9.3.

El procedimiento de trabajo con AUTODOCK es similar al descrito en el apartado 5.1.2 en lo referente a la preparación del ligando y la proteína y en los parámetros. Como tipos atómicos se considera el conjunto estándar. En el *docking* ciego el tamaño de la malla se fija a 181×181×181 y en el *docking* normal a 61×61×61. En cada ejecución, se requieren un total de 100 cónformeros (*docking* ciego) o 10 cónformeros (*docking* en la cavidad de unión). Se realizan sucesivas repeticiones para cada cálculo. Como modelos de la CXCR4 se consideran: MODEL\_CONGEN, MODEL\_MODELLER y el modelo sin *loops*.

En la predicción del sitio de unión se escoge para su análisis la conformación más próxima a la cavidad y no aquella de menor energía de *docking*. En las predicciones del modo de unión, se analizan tanto la conformación de mínima energía como aquella que mejor reproduce las hipótesis propuestas (apartado 9.3). Estrictamente, el criterio debería ser únicamente escoger aquel de menor energía de *docking*, ya que es el único criterio disponible cuando no se tienen datos experimentales de la cavidad de unión. No obstante, en este caso, se hace uso de dicha información. Para cada conformación, se determina la distancia de cada uno de los ocho nitrógenos del AMD3100 a cada uno de los oxígenos carboxílicos (O(sp<sup>3</sup>) y O(sp<sup>2</sup>)) de los tres residuos Asp<sup>171</sup>, Asp<sup>262</sup> y Glu<sup>288</sup>. Además, se analizan con el programa HBPLUS<sup>539</sup> la identificación de contactos por puente de hidrógeno.

En la Tabla 9.3 se resumen las distancias encontradas para los cónformeros más próximos al sitio de unión, expresadas como promedio y desviación estándar (en paréntesis) del total de repeticiones de cada *docking* ciego. Se observa cómo en el modelo sin *loops*, de las 6 repeticiones realizadas, en todos los casos entra algún compuesto a la cavidad de unión. Para MODEL\_CONGEN, en una de las primeras repeticiones se encuentra que un cónformero, que además coincide con el de menor energía de *docking*, sí que identifica la cavidad de unión. Por ello, se amplía el número de repeticiones hasta 7, no encontrándose en ningún otro caso un cónformero que entre en dicha cavidad. En MODEL\_MODELLER, ningún cónformero de cinco repeticiones identifica la cavidad de unión. En la Figura 9.10 se muestra la conformación final encontrada en el *docking* ciego frente a MODEL\_CONGEN. Se observa cómo el AMD3100 tiene más tendencia a unirse a Asp<sup>262</sup> (4.4 Å) y Glu<sup>288</sup> (3.9 Å), siguiendo el modelo

de interacción propuesto por Trent (apartado 9.3), aunque las distancias no corresponden estrictamente a distancias de enlace por puente de hidrógeno.

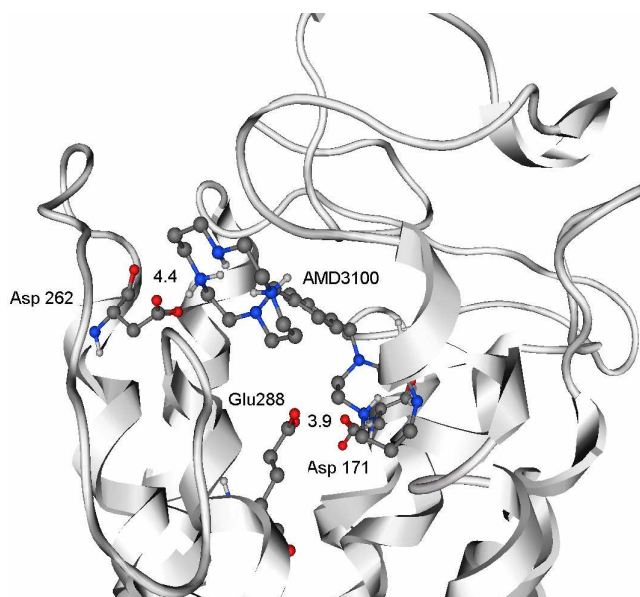
Dentro de los conformémeros en el *docking* frente a CXCR4 sin *loops*, uno de ellos presenta dos nitrógenos del anillo de ciclamó a distancias de puente de hidrógeno con Asp<sup>262</sup> (3.1 y 2.7 Å).

La mayor apertura del *loop* ECL2 en MODEL\_CONGEN frente a MODEL\_MODELLER (Figura 9.8) permite que en al menos en un caso el AMD3100 entre en la cavidad de unión. De hecho, otros autores<sup>527</sup> han apuntado las dificultades de reproducir el modo de unión de AMD3100(Zn<sub>2</sub>) en presencia de los *loops*, llegando incluso a sugerir que *in vivo* el proceso debería realizarse en condiciones reductoras que permitan la apertura del *loop* ECL2.

**Tabla 9.3.** *Docking* ciego de AMD3100 frente a los tres modelos de CXCR4 con AUTODOCK. En paréntesis se muestra la desviación estándar de las sucesivas repeticiones.

	Sin Loops	MODEL_MODELLER	MODEL_CONGEN
# Repeticiones	6	5	7
# Conformaciones en la cavidad de unión	6	0	1
O(sp <sup>3</sup> )Asp <sup>171</sup>	6.4 (0.7)	-	11.2
O(sp <sup>2</sup> )Asp <sup>171</sup>	5.5 (1.6)	-	8.1
O(sp <sup>3</sup> )Asp <sup>262</sup>	4.4 (1.1)*	-	4.4
O(sp <sup>2</sup> )Asp <sup>262</sup>	5.0 (1.2)*	-	6.5
O(sp <sup>3</sup> )Glu <sup>288</sup>	5.9 (1.6)	-	4.6
O(sp <sup>2</sup> )Glu <sup>288</sup>	6.3 (1.8)	-	3.9

\*para uno de los conformémeros se detectan puentes de hidrógeno con Asp<sup>262</sup> mediante el programa HBPLUS, con distancias de dos de sus nitrógenos de 3.1 y 2.7 Å.

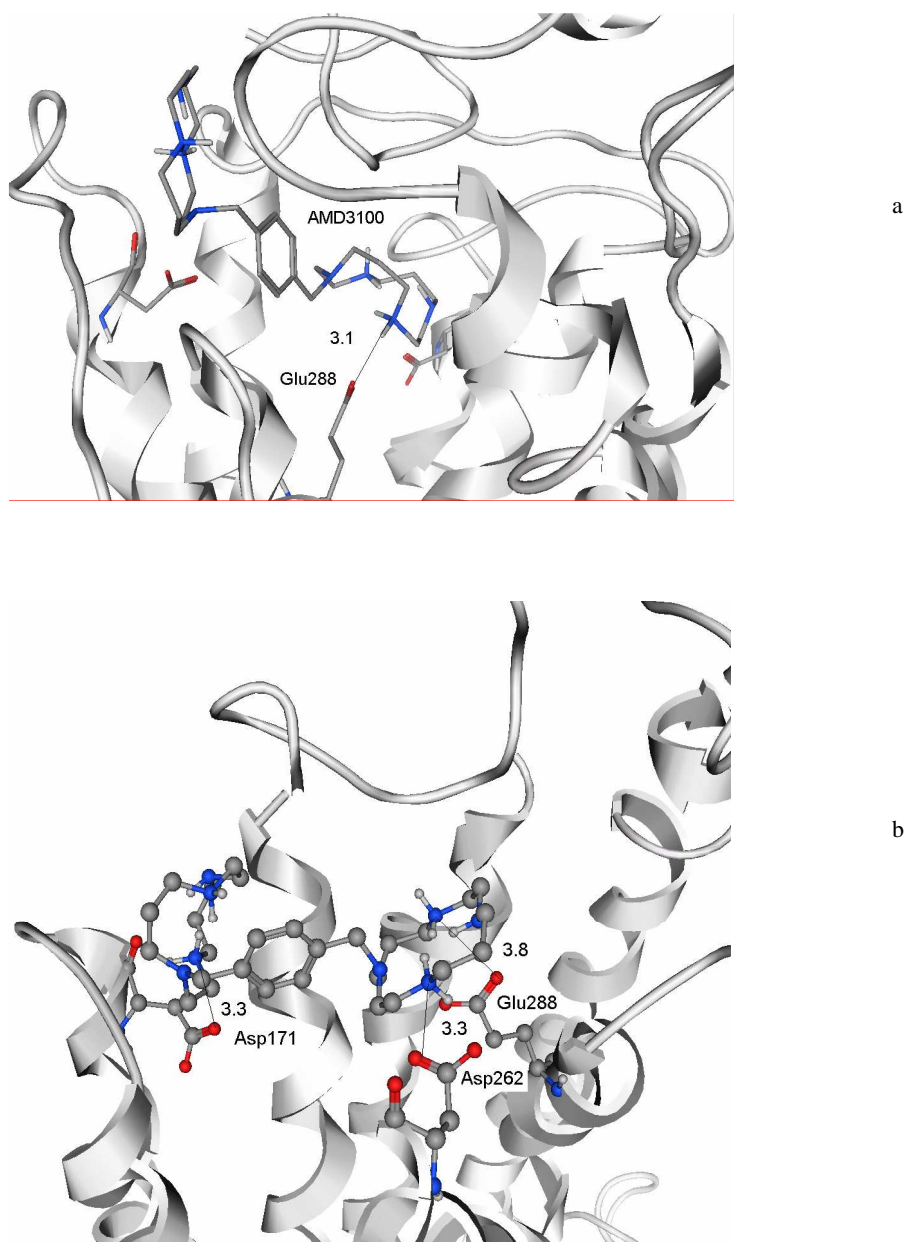


**Figura 9.10.** Conformación final del AMD3100 obtenida en el *docking* ciego frente a MODEL\_CONGEN en el que el conformémero de menor energía de *docking* identifica la cavidad de unión. Se indican las distancias frente a los dos residuos ácidos Asp<sup>262</sup> (4.4 Å) y Glu<sup>288</sup> (3.9 Å).

En la Tabla 9.4 se muestran los resultados obtenidos en la predicción del modo de unión, expresadas como la distancia promedio (desviación estándar en paréntesis) de las 5 repeticiones de cada *docking*. En general, los conformémeros de mínima energía de *docking* no corresponden a aquellos que mejor reproducen alguna de las hipótesis propuestas (que se sitúan en la segunda o tercera conformación de menor energía de *docking*).

Como en el caso anterior, frente a MODEL\_MODELLER el AMD3100 no se orienta adecuadamente en la cavidad de unión. Para MODEL\_CONGEN, las conformaciones resultantes tienden a acercarse más a Asp<sup>262</sup> y Glu<sup>288</sup>, presentando un conformémero a distancias de

punto de hidrógeno de Glu<sup>288</sup> (3.3 y 3.1 Å) (Figura 9.11.a). En el modelo sin *loops*, los conformeros se acercan a los tres residuos ácidos. De hecho, se detecta un conformero situado a distancias de 3.3, 3.8 y 3.3 Å de los residuos Asp<sup>171</sup>, Asp<sup>262</sup> y Glu<sup>288</sup>, respectivamente, correspondiendo al modelo “sándwich” descrito en el apartado 9.3 (Figura 9.11.b).



**Figura 9.11.** (a) *Docking* del AMD3100 frente MODEL\_CONGEN, donde el AMD3100 establece puente de hidrógeno con Glu288. (b) *Docking* del AMD3100 frente a CXCR4 sin *loops*, donde se orienta adoptando la conformación tipo “sándwich”.

**Tabla 9.4.** *Docking* de AMD3100 situado en la cavidad frente a los tres modelos de CXCR4 con AUTODOCK. *Mínima E*: corresponde a la conformación de mínima energía de *docking*. *Mejor reproducción*: corresponde a la conformación que mejor reproduce un modelo de modo de unión. En paréntesis se muestra la desviación estándar de las 5 repeticiones.

	<b>Sin Loops</b>		<b>MODEL_MODELER</b>		<b>MODEL_CONGEN</b>	
	Mínima E	Mejor reproducción	Mínima E	Mejor reproducción	Mínima E	Mejor reproducción
O(sp <sup>3</sup> )Asp <sup>171</sup>	6.5 (0.1)	6.4 (0.6)	22.3 (2.3)	23.1 (3.6)	11.3 (0.7)	10.8 (1.5)
O(sp <sup>2</sup> )Asp <sup>171</sup>	5.7 (0.7)	3.9 (0.8)*	23.1 (2.1)	22.4 (2.8)	9.1 (0.5)	9.3 (2.0)
O(sp <sup>3</sup> )Asp <sup>262</sup>	4.8 (0.4)	5.5(0.8)	10.3 (0.7)	7.2 (2.9)	5.3 (1.2)	4.7 (1.4)
O(sp <sup>2</sup> )Asp <sup>262</sup>	5.1 (0.4)	3.4 (0.3)*	9.0 (0.7)	6.4 (1.0)	6.3 (1.3)	6.0 (1.2)
O(sp <sup>3</sup> )Glu <sup>288</sup>	6.8 (0.6)	5.3 (1.1)	16.0 (3.8)	18.3 (3.2)	7.0 (1.5)	5.2 (2.0)*
O(sp <sup>2</sup> )Glu <sup>288</sup>	6.3 (0.6)	5.8 (3.0)*	15.1 (4.2)	15.6 (4.4)	4.7 (0.5)	4.4 (0.4)

\*para 2-3 de los conformeros se detectan puentes de hidrógeno de estos residuos mediante el programa HBPLUS. Concretamente, para CXCR4 sin *loops* uno de los conformeros se encuentra a distancias de 3.3, 3.8 y 3.3 Å de los Asp<sup>171</sup>, Asp<sup>262</sup> y Glu<sup>288</sup> (Figura 9.11.b).

Considerando estos resultados, la modelización del *loop* ECL2 con CONGEN permite identificar un modo de unión más consistente con los datos experimentales que el modelo obtenido con MODELLER. Sin embargo, dado que las distancias finales obtenidas con el *docking* no se aproximan lo suficiente como para reproducir puentes de hidrógeno, está pendiente de realizarse en un futuro estudios de dinámica molecular a partir de las conformaciones finales resultantes del *docking* que permitan muestrear la cavidad de unión con un modelo más refinado que el *docking*.

La situación sin *loops* es la que permite obtener unos resultados más similares a la hipótesis de una interacción tanto con Asp<sup>262</sup> como con Asp<sup>171</sup>. El interés de incluir los *loops* en el modelo reside en la aplicación posterior del modelo de la CXCR4 en el cribado virtual. Existen estudios<sup>540</sup> que apuntan a que un cribado virtual que incluye los *loops* de los modelos de GPCRs, particularmente ECL2, discrimina mejor entre compuestos activos e inactivos. La inspección visual de las orientaciones de los compuestos en estos estudios muestra que la eliminación del *loop* ECL2 provoca que los compuestos inactivos más flexibles se orienten artificialmente a la región libre del *loop* ECL2, resultando en mejores *scores* e incrementando el número de falsos positivos.

## 9.7. Estudio del sitio y modo de unión del TAK-779 en CCR5

Como en el caso de AMD3100 frente a CXCR4, se realiza el mismo estudio del modo de unión de TAK-779 frente CCR5 (apartado 9.4).

Se considera un pH de trabajo de 7, manteniéndose los residuos ácidos desprotonados y los básicos protonados. Como en el caso de CXCR4, se analizan con AUTODOCK:

- Predicción del sitio de unión del TAK-779 en CCR5 mediante *docking* ciego.
- Predicción del modo de unión mediante *docking*, en el que el TAK-779 se sitúa inicialmente en la cavidad de unión descrita en el apartado 9.4.

La preparación del ligando/proteína y los parámetros de trabajo corresponden a los descritos en el apartado 9.6. En este caso se trabaja sobre la estructura MODEL\_MODELER y la estructura sin *loops* de CCR5. Para las conformaciones resultantes, se evalúa la distancia del nitrógeno cuaternario a cada uno de los oxígenos carboxílicos de Glu<sup>283</sup>, así como las interacciones hidrofóbicas (HBPLUS) entre TAK-779 y los residuos Tyr<sup>37</sup>, Trp<sup>86</sup>, Tyr<sup>108</sup>, Leu<sup>33</sup>, Arg<sup>31</sup>, Ile<sup>198</sup> y Thr<sup>82</sup>. Como en el caso anterior, en el *docking* ciego se selecciona la conformación más próxima



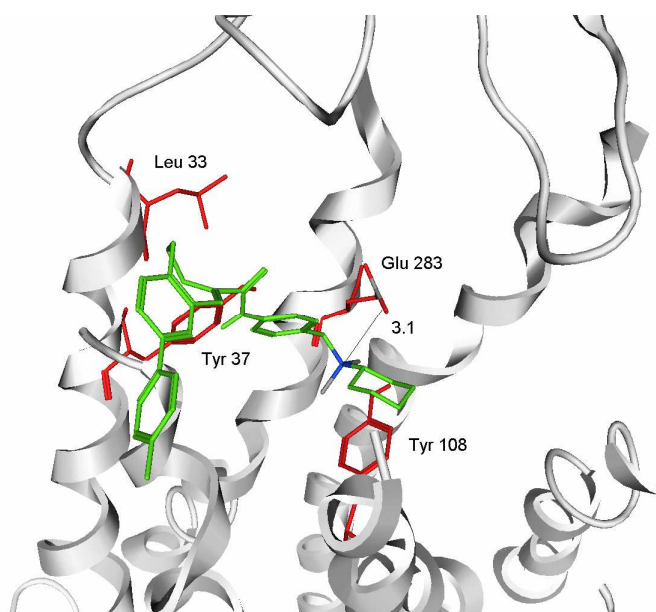
a la cavidad de unión y en la predicción del modo de unión, tanto la conformación de mínima energía de *docking* como la que mejor se ajusta a los datos experimentales.

En la Tabla 9.5 se resumen las distancias encontradas para los conformeros más próximos al sitio de unión, expresadas como promedio y desviación estándar (en paréntesis) del total de repeticiones de cada *docking* ciego. En negrita se indican aquellos residuos cuya mutación tiene más efecto en la actividad antiviral del TAK-779. Junto a la desviación estándar, se muestra en cursiva el número de repetición que identifica el contacto mostrado. Como en el caso de la CXCR4, la inclusión de los *loops* modelados con MODELLER impide la entrada de TAK-779 en la cavidad de unión. Para CCR5 sin *loops*, los conformeros se sitúan en la cavidad de unión. En particular, una de las 6 repeticiones identifica un conformero que establece la interacción TAK-779 con el Glu<sup>283</sup> (3.1 Å) (Figura 9.12), el resto de repeticiones se encuentran a distancias comprendidas entre 3.9 y 5.2 Å. Las interacciones hidrofóbicas con aquellos residuos más importantes (Tyr<sup>37</sup>, Trp<sup>86</sup>, Tyr<sup>108</sup>) se encuentran también para el conformero que establece la interacción con Glu<sup>283</sup> (identificado como 2 en la Tabla 9.5, mostrado en la Figura 9.12).

**Tabla 9.5.** *Docking* ciego de TAK-779 frente a los dos modelos de CCR5 con AUTODOCK. En paréntesis se muestra la desviación estándar de las sucesivas repeticiones.

	<b>Sin Loops</b>	<b>MODEL MODELLER</b>
# Repeticiones	6	5
# Conformaciones en la cavidad de unión	6	0
O(sp <sup>3</sup> )Glu <sup>283</sup>	4.3 (0.3)	11.4 (1.7)
O(sp <sup>2</sup> )Glu <sup>283</sup>	4.0 (0.6)*(2)	12.8 (0.9)
<b>Tyr<sup>37</sup></b>	3.1 (0.2) (2,4,6)	
<b>Trp<sup>86</sup></b>	3.0 (0.3) (1,2,5)	
<b>Tyr<sup>108</sup></b>	2.7 (0.1) (2,6)	
<b>Leu<sup>33</sup></b>		3.0 (0.1) (1,2,4)
Arg <sup>31</sup>	3.1 (0.0) (6)	
Ile <sup>198</sup>	2.8 (0.0) (3)	
Thr <sup>82</sup>		

\*para 1 conformero se detecta interacción con Glu<sup>283</sup> (Figura 9.12). Los números en *cursiva* indican la identidad del conformero (número de repetición del *docking*) para la que se encuentran las interacciones descritas en la tabla.



**Figura 9.12.** *Docking* ciego del TAK-779 (verde) frente CCR5 sin *loops*. TAK-779 establece interacción con Glu<sup>283</sup> (3.1 Å) así como con los residuos Tyr<sup>37</sup> (distancia de interacción hidrofóbica de 3.2 Å), Trp<sup>86</sup> (distancia de interacción hidrofóbica de 2.9 Å) y Tyr<sup>108</sup> (distancia de interacción hidrofóbica de 2.6 Å).

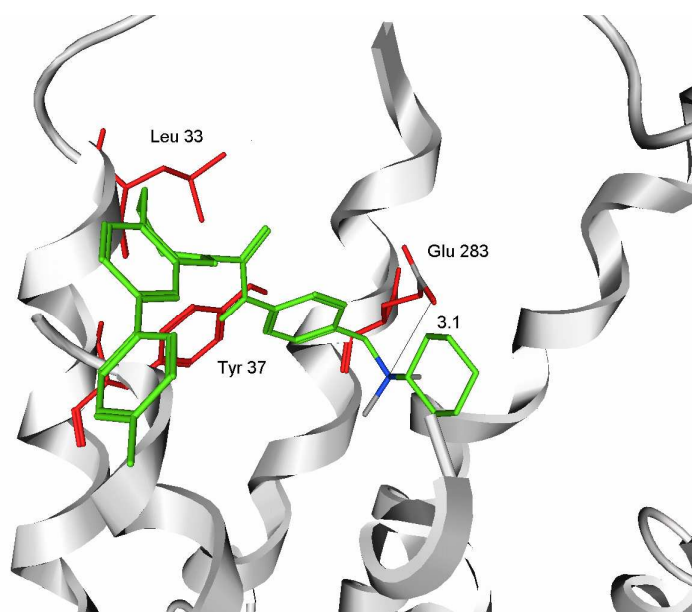


En la Tabla 9.6 se muestran los resultados obtenidos en la predicción del modo de unión, expresadas como la distancia promedio (desviación estándar en paréntesis) de las 5 repeticiones de cada *docking* en la cavidad de unión. El *docking* frente a MODEL\_MODELLEER resulta de nuevo fallido en la identificación de la orientación adecuada. En el modelo sin *loops*, al seleccionar aquellos conforméromos que mejor identifican el modelo propuesto, se obtienen en las cinco repeticiones orientaciones con interacción entre el nitrógeno amónico y Glu<sup>283</sup> (con distancias comprendidas entre 3.0 y 3.9 Å). También se detectan las interacciones hidrofóbicas citadas. Interesantemente, en caso de analizar la conformación de menor energía de *docking*, se detectan en todos los casos interacciones por puente de hidrógeno entre el nitrógeno peptídico de TAK-779 y Glu<sup>283</sup> (con distancias comprendidas entre 2.6-3.4 Å).

**Tabla 9.6.** *Docking* TAK-779 situado en la cavidad frente a los dos modelos de CCR5 con AUTODOCK. *Mínima E*: corresponde a la conformación de mínima energía de *docking*. *Mejor reproducción*: corresponde a la conformación que mejor reproduce un modelo de modo de unión. En paréntesis se muestra la desviación estándar de las 5 repeticiones.

	<b>Sin Loops</b>		<b>MODEL_MODELLEER</b>	
	Mínima E	Mejor reproducción	Mínima E	Mejor reproducción
O(sp <sup>3</sup> )Glu <sup>283</sup> N amónico	9.9 (0.2)	4.4 (0.3)	14.4 (0.7)	10.3 (0.3)
O(sp <sup>2</sup> )Glu <sup>283</sup> N amónico	8.2 (0.1)	3.5 (0.3)	16.0 (0.9)	11.5 (0.6)
O(sp <sup>3</sup> )Glu <sup>283</sup> N peptídico	5.3 (0.3)	7.7 (0.5)		13.7 (0.6)
O(sp <sup>2</sup> )Glu <sup>283</sup> N peptídico	3.1 (0.3)	8.3 (0.8)		14.5 (1.6)
Tyr <sup>37</sup>	2.9 (0.1)	3.2 (0.4)		
Trp <sup>86</sup>		3.1 (0) (5)		
Tyr <sup>108</sup>	3.3 (0.1)	3.1 (0.1) (3,4,5)		
Leu <sup>33</sup>			3.55 (0) (1)	
Arg <sup>31</sup>		3.2 (0.2) (2,3,4)	3.1 (0.1) (2,3,4,5)	
Ile <sup>198</sup>	3.3 (0.2)			
Thr <sup>82</sup>				

En la Figura 9.13 se muestra la conformación resultante del *docking* de TAK-779 (*verde*) frente CCR5 sin *loops*, en la que se establece un puente de hidrógeno entre el nitrógeno amónico y el Glu<sup>283</sup> a distancia de 3.1 Å. Además este conforméromo interactúa con Tyr<sup>37</sup> (distancia hidrofóbica de 3.6 Å), Trp<sup>86</sup> (distancia hidrofóbica de 3.1 Å) y Tyr<sup>108</sup> (distancia hidrofóbica de 3.2 Å).



**Figura 9.13.** *Docking* del TAK-779 (*verde*) frente CCR5 sin *loops*. TAK-779 establece interacción con Glu<sup>283</sup> (3.1 Å) así como con los residuos Tyr<sup>37</sup> (distancia de interacción hidrofóbica de 3.6 Å), Trp<sup>86</sup> (distancia de interacción hidrofóbica de 3.1 Å) y Tyr<sup>108</sup> (distancia de interacción hidrofóbica de 3.2 Å).

Dado el impacto de los *loops*, en los que MODEL\_MODELLEER no consigue reproducir el modo de unión propuesto experimentalmente, se propone su re-modelización con CONGEN en un futuro, con el fin de observar si se repite el comportamiento observado para CXCR4.

Los modelos de CXCR4 y CCR5 están siendo utilizados en un estudio de cribado virtual. El ensayo retrospectivo se realiza para una base de datos de aproximadamente 4700 compuestos, de los cuales 248 y 354 son inhibidores de CXCR4 y CCR5, respectivamente, con quimiotipos representativos de los antagonistas descritos en los apartados 9.1 y 9.2. Los primeros resultados obtenidos con la función de AUTODOCK alcanzan factores de enriquecimiento del orden de 2-3 en el primer 10% de base de datos cribada. Estos modelos se están utilizando también en la evaluación de una serie de 88 compuestos de estructuras amina-*linker*-amina, amina-*linker*-aldehído, ciclamo-*linker*-amina, sintetizados en el laboratorio de síntesis del IQS<sup>541</sup>, uno de los cuales ha mostrado actividad en el orden del AMD3100 (0.001µg/mL). Dado que todos estos estudios se encuentran en este momento en una fase inicial, no se discuten en la presente tesis, siendo objeto de estudios futuros.

## Conclusiones

La presente tesis ha tratado de establecer y validar un protocolo de cribado virtual para la identificación de inhibidores de receptores de tirosina quinasas, particularmente FGFR, EGFR y PDGFR. Se recopilan algunas de las conclusiones, perfiladas en los capítulos respectivos, principalmente en el contexto de la aplicación retrospectiva (validación) ya que no se dispone actualmente de datos de actividad para los compuestos sintetizados que permitan evaluar de una manera prospectiva las diferentes etapas.

- Se ha diseñado una quimioteca virtual de pirido[2,3-*d*]pirimidinas formada por 106920 compuestos sobre la que se han realizado selecciones combinatorias independientes en función de las distintas sustituciones (NH<sub>2</sub>, OH, H) en la posición del carbono cuatro. Aparte de la finalidad sintética de la selección, se ha analizado el comportamiento de los métodos de partición, encontrándose que la función *cell-based Chi2* es la que presenta un compromiso óptimo de recubrimiento en términos de espacio y población, al menos para dicha quimioteca. La pérdida de recubrimiento originada por el tratamiento independiente de cada una de las tres quimiotecas se supera mediante una optimización multiobjetivo que reúna todas las funciones en un único criterio a optimizar.
- Los tres métodos de búsqueda farmacofórica (CATS3D, SQUID y el modelo farmacofórico del MOE) permiten capturar un número significativo de compuestos activos. El mayor enriquecimiento se obtiene con el modelo refinado manualmente de MOE. Al eliminar los puntos potenciales farmacofóricos con múltiple asignación de tipos en el modelo obtenido con MOE, SQUID es el método que permite obtener mayores factores de enriquecimiento. Además de determinar los factores de enriquecimiento, se ha evaluado la diversidad en la recuperación de *scaffolds*. Desde este punto de vista, tanto MOE como SQUID permiten la identificación de varios quimiotipos en los primeros porcentajes de base de datos cribada, complementándose ambos. CATS3D, si bien opera satisfactoriamente desde el punto de vista de enriquecimiento, no explora tan bien como los anteriores la diversidad de los *scaffolds*. Así, se aconseja el uso de los modelos farmacofóricos de MOE y SQUID en la evaluación prospectiva.
- Se ha modificado la implementación original de SQUID, introduciéndose la conexión modelo farmacofórico-*fingerprint* SQUID-SQUID *not scaled* que permite aumentar la eficiencia al prescindir de una etapa de optimización de pesos así como habilitar el uso de las métricas tradicionalmente usadas en las búsquedas de similitud. En el caso de los inhibidores de tirosina quinasa, esta conexión aumenta considerablemente el factor de enriquecimiento respecto a la conexión SQUID-CATS3D, aunque no se ha constatado el mismo aumento para casos como el de los inhibidores de COX-2, por lo que no puede asegurarse que una conexión supere a otra en términos de eficacia.
- La reproducción del modo de unión de una serie de inhibidores de tirosina quinasa mediante *docking* es altamente dependiente de la elección de un sitio de unión apropiado, afectando por ello a la recuperación de quimiotipos diversos en un cribado virtual. Tanto GOLD como AUTODOCK son capaces de predecir el modo de unión de inhibidores co-cristalizados con la estructura rígida utilizada en el *docking*, aunque la tasa de éxito disminuye al realizar experimentos de *cross-docking*. Además, AUTODOCK permite la identificación del sitio y modo de unión de una serie de inhibidores de FGFR, EGFR y PDGFR mediante *docking* ciego. El cribado virtual con AUTODOCK y GOLD permite discriminar entre activos e inactivos por encima de una selección aleatoria, particularmente en el caso de FGFR. Para esta quinasa, la estructura 2fgi co-cristalizada con PD173074 es la más apropiada, tanto en términos de enriquecimiento como de recuperación de quimiotipos diversos.

En el caso de EGFR, únicamente AUTODOCK recupera activos por encima de una selección aleatoria. El modelo por homología de PDGFR es válido para realizar cribados virtuales, corroborándose así el uso de tales modelos para esta familia de proteína quinasas en estudios de *docking*. Junto con la influencia de la estructura, la introducción de restricciones de interacción o el *rescoring* resultan beneficiosas para el cribado virtual. Además, el uso de estrategias de *consensus scoring* permite aprovechar la información contenida en varias estructuras de una misma proteína.

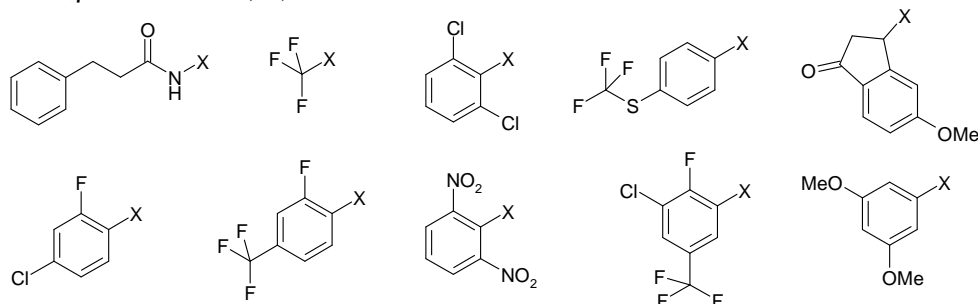
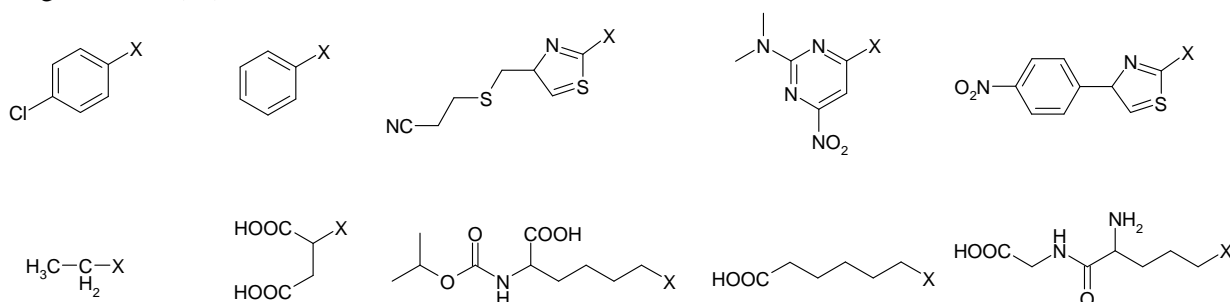
- El cribado virtual retrospectivo con una aproximación indirecta (*ligand-based*) y directa (*structure-based* frente FGFR) alcanza enriquecimientos similares en los primeros porcentajes de base de datos cribada. Desde un punto de vista de la diversidad de los quimiotipos, existe una complementación deseable entre ambas aproximaciones.
- Se ha diseñado un *fingerprint* de interacción estructural basado en pares de átomos (IFbAP), cuya aplicación en una búsqueda de similitud permite recuperar compuestos con factores de enriquecimiento superiores a una selección aleatoria. El uso del criterio combinado de similitud con el criterio energético (*docking*) permite incrementar la recuperación de activos, obviamente excepto si la propia función de *scoring* se comporta mal. Comparativamente frente a otro *fingerprint* estructural como CHIF, se ha encontrado que en algunos casos CHIF supera a IFbAP (receptor de estrógeno) y en otros casos IFbAP opera mejor (transcriptasa reversa). Este *fingerprint*, así como sus análogos, son fácilmente implementables, requiriendo un coste computacional muy bajo.
- La aplicación del método MM-PBSA/MM-GBSA para la predicción de afinidad de un total de 27 inhibidores de FGFR permite alcanzar factores de correlación de  $R^2$  de 0.51 al realizarse las dinámicas en condiciones de solvente explícito. Sin embargo, en otras condiciones de dinámica, los coeficientes de correlación decaen situándose en valores similares a los obtenidos con funciones de *scoring*. A nuestro entender, la diferencia de coste computacional asociado no justifica totalmente el empleo de MM-PBSA, al menos para el caso de estudio y las condiciones aplicadas. La aplicación del método MM-GBSA en cribado virtual ha resultado positiva, superando el enriquecimiento obtenido con AUTODOCK, aunque de nuevo, a un mayor coste computacional.
- Se han habilitado en PRALINS diversos algoritmos para la selección focalizada de quimiotecas combinatorias. El método Direct propuesto representa un buen compromiso entre la habilidad para focalizar, la velocidad y su facilidad de implementación.
- Se ha definido una medida de diversidad, *cell-integral-diversity criterion*, que permite evaluar el recubrimiento de las selecciones con una alta independencia de la partición de referencia utilizada, evitándose los artefactos del *cluster* asociados a los métodos *cell-based*. Además, disminuye la tendencia al *edge design* asociada a métodos basados en distancia, como la evaluación *diversity integral criterion*. Por otra parte, su aplicación como función objetivo para seleccionar *full arrays* permite una distribución más homogénea de los compuestos respecto a la obtenida con la función *cell-based Fraction*.
- Se ha implementado la optimización multiobjetivo en PRALINS, permitiéndose la selección de compuestos que optimizan la diversidad, el tamaño y configuración de los *full arrays* así como otras propiedades. La comparación frente a la obtención de soluciones individuales mediante optimizaciones con algoritmos genéticos o *simulated annealing* permite concluir que MOGA identifica conjuntos óptimos de Pareto a un mismo nivel de eficacia, aunque con una considerable mayor eficiencia. Por otra parte, el uso de algoritmos genéticos en PRALINS como técnica de optimización es en general más eficaz que el *simulated annealing*, aunque con un mayor requerimiento computacional.

- Los modelos transmembrana de CXCR4 y CCR5 son consistentes con las evidencias experimentales, resultando útiles para el estudio del modo de unión de dos de sus antagonistas. La inclusión de los *loops* obtenidos con MODELLER deteriora los resultados obtenidos con los modelos sin *loops*, aunque el refinamiento de dichos segmentos para CXCR4 mediante CONGEN ha permitido la descripción del modo de unión experimental propuesto.

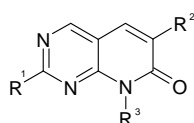


## Anexo

**Anexo 1.** Listado de reactivos ésteres  $\alpha,\beta$ -insaturados ( $R^4$ ) y guanidinas ( $R^1$ ) seleccionados finalmente para ser sintetizadas en una quimioteca combinatoria 10×10 (Capítulo 3). El *dummy atom* X indica el punto de anclaje a la molécula.

10 ésteres  $\alpha,\beta$ -insaturados ( $R^4$ )10 guanidinas ( $R^1$ )

**Anexo 2.** Listado de moléculas inhibidoras de FGFR, EGFR y PDGFR que componen el *pool* de activos en las respectivas validaciones retrospectivas.

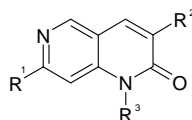
Pirido[2,3-*d*]pirimidinas:

Nº	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	IC <sub>50</sub> ( $\mu$ M) PDGFR	IC <sub>50</sub> ( $\mu$ M) FGFR	IC <sub>50</sub> ( $\mu$ M) 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

Nº	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	IC <sub>50</sub> (µM) PDGFR	IC <sub>50</sub> (µM) FGFR	IC <sub>50</sub> (µM) EGFR
18	NH(CH <sub>2</sub> ) <sub>5</sub> COOH	2,6-diClPhe	Me	0.73	0.58	1.0
19	NHPh	2,6-diClPhe	Me	0.40	0.46	0.26
20	NHPh	2,6-diClPhe	Et	0.80	0.21	0.20
21	NH-3-piridil	2,6-diClPhe	Me	0.12	0.15	0.51
22	NH-3-piridil(6-OMe)	2,6-diClPhe	Me	0.40	0.54	1.3
23	NH-4-piridil	2,6-diClPhe	Me	0.10	0.19	0.91
24	NH-4-piridil	2,6-diClPhe	Et	0.10	0.11	0.49
25	NHPh(4-Cl)	2,6-diClPhe	Me	2.6	3.0	>5
26	NHPh(3-Br)	2,6-diClPhe	Me	3.1	2.4	0.67
27	NHPh(4-Me)	2,6-diClPhe	Me	1.2	0.91	0.48
28	NHPh(2-OMe)	2,6-diClPhe	Me	1.8	3.2	0.66
29	NHPh(3-OMe)	2,6-diClPhe	Me	0.53	0.37	0.03
30	NHPh(4-OMe)	2,6-diClPhe	Me	0.68	0.41	0.22
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
34	NHPh(3-Me,4-OMe)	2,6-diClPhe	Me	1.5	0.45	0.25
35	NHPh(3,5-(OMe) <sub>2</sub> )	2,6-diClPhe	Me	1.7	0.43	0.33
36	NHPh(3-COOEt)	2,6-diClPhe	Me	1.7	1.5	nt
37	NHPh(4-COOEt)	2,6-diClPhe	Me	21.4	33.6	>50
38	NHPh(4-(CH <sub>2</sub> ) <sub>3</sub> CO <sub>2</sub> Et)	2,6-diClPhe	Me	4.2	2.6	0.70
39	NHPh(4-O(CH <sub>2</sub> ) <sub>2</sub> NH <sub>2</sub> )	2,6-diClPhe	Me	0.074	0.055	nt
40	NHPh(4-O(CH <sub>2</sub> ) <sub>2</sub> NHEt)	2,6-diClPhe	Me	0.19	0.066	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>2</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
46	NHPh(4-O(CH <sub>2</sub> ) <sub>2</sub> N <sup>+</sup> (Me)Et <sub>2</sub> I)	2,6-diClPhe	Me	0.081	0.040	nt
47	NHPh(4-O(CH <sub>2</sub> ) <sub>2</sub> N(O)Et <sub>2</sub> )	2,6-diClPhe	Me	0.12	0.046	nt
48	NHPh(4-O(CH <sub>2</sub> ) <sub>2</sub> Npiper <sup>c</sup> )	2,6-diClPhe	Me	0.098	0.039	nt
49	NHPh(4-Npip(Me) <sup>b</sup> )	2,6-diClPhe	Me	0.11	0.028	nt
50	NHPh(4-CH <sub>2</sub> NpipMe) <sup>b</sup>	2,6-diClPhe	Me	0.101	0.049	nt
51	NHPh(3-COOH)	2,6-diClPhe	Me	0.13	0.11	0.19
52	NHPh(4-CH <sub>2</sub> CO <sub>2</sub> H)	2,6-diClPhe	Me	0.072	0.061	0.22
53	NHPh(4-(CH <sub>2</sub> ) <sub>3</sub> CO <sub>2</sub> H)	2,6-diClPhe	Me	0.12	0.11	0.08
54	NH <sub>2</sub>	2,6-diClPhe	Et	0.98	0.54	nt
55	NH <sub>2</sub>	2,6-diClPhe	<i>n</i> -Pr	1.2	0.51	nt
56	NH <sub>2</sub>	2,6-diClPhe	<i>n</i> -Bu	1.2	0.58	nt
57	NH <sub>2</sub>	2,6-diClPhe	<i>i</i> -Bu	1.6	0.50	nt
58	NH <sub>2</sub>	2,6-diClPhe	CH <sub>2</sub> COOMe	5.0	1.7	nt
59	NH <sub>2</sub>	2,6-diClPhe	CH <sub>2</sub> COO- <i>t</i> -Bu	7.2	1.7	nt
60	NH <sub>2</sub>	2,6-diClPhe	CH <sub>2</sub> COOH	>50	7.9	nt
61	NH <sub>2</sub>	2,6-diClPhe	CH <sub>2</sub> Ph	4.2	1.4	nt
62	NH <sub>2</sub>	2,6-diClPhe	CH <sub>2</sub> -4piridinil	1.2	0.72	nt
63	NH <sub>2</sub>	2,6-diClPhe	(CH <sub>2</sub> ) <sub>3</sub> NMe <sub>2</sub>	1.8	0.88	nt
64	NH <sub>2</sub>	2,6-diClPhe	(CH <sub>2</sub> ) <sub>3</sub> OCH <sub>2</sub> Ph	6.8	3.2	nt
65	NH <sub>2</sub>	2,6-diClPhe	(CH <sub>2</sub> ) <sub>3</sub> OH	1.1	0.37	nt
66	NHPh	Ph	Me	0.29	5.0	nt
67	NHPh	3-tiofeno	Me	0.80	>50	nt
68	NH-4-piridina	Ph	Me	0.32	4.7	nt
69	NH-4-piridina	3-tiofeno	Me	0.48	22.6	nt
70	NH-Ph-4-O(CH <sub>2</sub> ) <sub>2</sub> NEt <sub>2</sub>	Ph	Me	0.042	0.35	nt
71	NH-Ph-4-O(CH <sub>2</sub> ) <sub>2</sub> NEt <sub>2</sub>	3-tiofeno	Me	0.054	2.9	nt
72	NH <sub>2</sub>	3-tiofeno	Me	24.4	>50	nt
73	NHPh	2-tiofeno	Me	0.95	>50	nt
74	NHPh	Ph	Et	0.18	1.5	nt
75	NHPh	3-tiofeno	Et	0.41	>50	nt
76	NH-4-piridina	Ph	Et	0.074	1.1	nt
77	NH-Ph-4-O(CH <sub>2</sub> ) <sub>2</sub> NEt <sub>2</sub>	Ph	Et	0.031	0.088	nt
78	NHPh(morph)	2,6-diClPhe	Me	0.64	0.10	nt
79	NHPh(4-CONEt <sub>2</sub> )	2,6-diClPhe	Me	0.29	0.18	nt

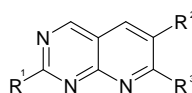
<sup>a</sup>Morfo:morfolin-1-il. <sup>b</sup>pip:N-metilpiperazin-1-il. <sup>c</sup>piper:piperidin-1-il. nt: no testado.



Naftiridin-2(1*H*)-onas:

Nº	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	IC <sub>50</sub> (µM) PDGFR	IC <sub>50</sub> (µM) FGFR
1	NH <sub>2</sub>	2,6-diClPhe	Me	3.6	0.38
2	NHMe	2,6-diClPhe	Me	8.0	0.21
3	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-diClPhe	Me	4.6	0.15
4	NH(CH <sub>2</sub> ) <sub>4</sub> NEt <sub>2</sub>	2,6-diClPhe	Me	2.4	0.080
5	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-diClPhe	Me	0.74	0.080
6	NH(CH <sub>2</sub> ) <sub>3</sub> (4-Me-pip) <sup>b</sup>	2,6-diClPhe	Me	2.8	0.17
7	NH(CH <sub>2</sub> ) <sub>4</sub> (4-Me-pip) <sup>b</sup>	2,6-diClPhe	Me	1.1	0.17
8	NH(CH <sub>2</sub> ) <sub>3</sub> (4-Me-pip) <sup>b</sup>	2,6-diClPhe	Me	9.9	0.11
9	NH(CH <sub>2</sub> ) <sub>3</sub> (morph) <sup>c</sup>	2,6-diClPhe	Me	5.8	0.57
10	NH(CH <sub>2</sub> ) <sub>4</sub> (morph) <sup>c</sup>	2,6-diClPhe	Me	3.1	0.23
11	NH(CH <sub>2</sub> ) <sub>3</sub> (1-imid) <sup>d</sup>	2,6-diClPhe	Me	3.2	0.21
12	MHPPh	2,6-diClPhe	Me	1.4	2.2
13	NH(4-piridinil)	2,6-diClPhe	Me	6.9	3.5
14	NHPhOMe	2,6-diClPhe	Me	6.8	2.2
15	NHPhO(CH <sub>2</sub> ) <sub>2</sub> NEt <sub>2</sub> <sup>e</sup>	2,6-diClPhe	Me	0.10	0.13
16	NHPhO(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub> <sup>e</sup>	2,6-diClPhe	Me	0.36	0.25
17	NHPhO(CH <sub>2</sub> ) <sub>2</sub> (4-Me-pip) <sub>2</sub> <sup>b,e</sup>	2,6-diClPhe	Me	0.26	0.21
18	NHPhO(CH <sub>2</sub> ) <sub>3</sub> (4-Me-pip) <sub>2</sub> <sup>b,e</sup>	2,6-diClPhe	Me	0.90	0.12
19	NHPh(4-Me-pip) <sup>b,e</sup>	2,6-diClPhe	Me	0.26	0.042
20	NHPh(morph) <sup>c,e</sup>	2,6-diClPhe	Me	2.8	8.7
21	NHPh(4-CONEt <sub>2</sub> )	2,6-diClPhe	Me	16	1.9
22	NH(CH <sub>2</sub> ) <sub>4</sub> NEt <sub>2</sub>	2,6-diClPhe	H	1.1	0.17
23	NH(CH <sub>2</sub> ) <sub>3</sub> (4-Me-pip) <sup>b</sup>	2,6-diClPhe	H	1.4	0.072
24	NHPhO(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub> <sup>e</sup>	2,6-diClPhe	H	0.79	0.44

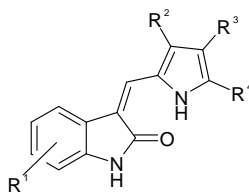
<sup>b</sup>4-Metilpiperazin-1-il. <sup>c</sup>N-Morfolinil. <sup>d</sup>Imidazolil. <sup>e</sup>anillo fenilo *para*-sustituido.

7-alkilurea pirido[2,3-*d*]pirimidinas:

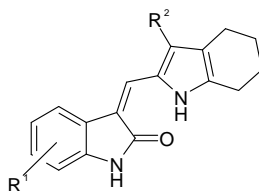
Nº	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	IC <sub>50</sub> (µM) PDGFR	IC <sub>50</sub> (µM) FGFR	IC <sub>50</sub> (µM) EGFR
1	H	Ph	NHCONH( <i>t</i> -Bu)	4.67	3.71	5.53
2	H	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	1.11	0.13	0.45
3	H	2,6-(Me) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	0.34	0.4	0.61
4	H	2,3,5,6-(Me) <sub>4</sub> Ph	NHCONH( <i>t</i> -Bu)	>50	0.78	6.68
5	H	3,5-(OMe) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	>50	0.060	>50
6	H	2,6-(Cl) <sub>2</sub> Ph	NHCONH(Et)	1.3	0.13	1.36
7	(CH <sub>2</sub> ) <sub>3</sub> N(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> NMe	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	0.66	0.082	6.20
8	(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	0.47	0.051	0.15
9	(CH <sub>2</sub> ) <sub>4</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	0.31	0.048	0.24
10	(CH <sub>2</sub> ) <sub>4</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH(Et)	0.19	0.033	1.26
11	(CH <sub>2</sub> ) <sub>3</sub> N(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> NMe	2,3,5,6-(Me) <sub>4</sub> Ph	NHCONH( <i>t</i> -Bu)	>50	0.14	7.0
12	H	2,3,6-(Cl) <sub>3</sub> Ph	NHCONH( <i>t</i> -Bu)	2.96	0.11	
13	H	2,6-(Br) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	1.42	0.29	
14	H	2-Br-6-ClPh	NHCONH( <i>t</i> -Bu)	0.62	0.18	
15	H	2,6-(F) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	1.67	0.11	
16	H	3,5-(F) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	8.97	1.10	
17	H	2-MePh	NHCONH( <i>t</i> -Bu)	1.05	1.40	
18	H	4-MePh	NHCONH( <i>t</i> -Bu)	6.31	1.67	
19	H	2,3-(Me) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	6.05	0.34	
20	H	3,5-(Me) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	52.90	1.13	
21	H	2,4,6-(Me) <sub>3</sub> Ph	NHCONH( <i>t</i> -Bu)	1.47	0.27	

N°	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	IC <sub>50</sub> (μM) PDGFR	IC <sub>50</sub> (μM) FGFR
22	H	2,3,4,5,6-(Me) <sub>5</sub> Ph	NHCONH( <i>t</i> -Bu)	>50	1.62
23	H	2,6-(Me) <sub>2</sub> ,3-OCH <sub>2</sub> CH <sub>2</sub> NEt <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	11.81	0.89
24	H	2-EtPh	NHCONH( <i>t</i> -Bu)	4.48	11.22
25	H	3,5-(Et) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	>50	7.76
26	H	2-OMePh	NHCONH( <i>t</i> -Bu)	4.48	11.22
27	H	3-OMePh	NHCONH( <i>t</i> -Bu)	22.93	0.36
28	H	4-OMePh	NHCONH( <i>t</i> -Bu)	2.89	3.97
29	H	3,4-(OMe) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	>50	20.25
30	H	3-OEt	NHCONH( <i>t</i> -Bu)	23.06	0.67
31	H	3,5-(OEt) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	>50	1.65
32	H	3,5-(NMe <sub>2</sub> ) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	>50	16.00
33	NH <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NH <sub>2</sub>	16	3.0
34	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NH <sub>2</sub>	46	2.4
35	NH <sub>2</sub>	2,6-(Me) <sub>3</sub> Ph	NH <sub>2</sub>	29	13
36	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Me) <sub>3</sub> Ph	NH <sub>2</sub>	25	18
37	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Me) <sub>3</sub> Ph	NHCONH( <i>t</i> -Bu)	0.80	0.34
38	NH <sub>2</sub>	2,6-(Br) <sub>2</sub> Ph	NH <sub>2</sub>	>50	13
39	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Br) <sub>2</sub> Ph	NH <sub>2</sub>	>50	8.3
40	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Br) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	1.1	0.19
41	NH <sub>2</sub>	3,5-(MeO) <sub>2</sub> Ph	NH <sub>2</sub>	>50	0.23
42	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH(Et)	12	1.3
43	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	9.0	1.8
44	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH(Et)	1.3	0.13
45	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>i</i> -Pr)	1.1	0.077
46	NH(CH <sub>2</sub> ) <sub>4</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH(Et)	0.21	0.049
47	NH(CH <sub>2</sub> ) <sub>4</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	0.36	0.048
48	NH(CH <sub>2</sub> ) <sub>4</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH(ciclohexil)	0.33	0.043
49	NH(CH <sub>2</sub> ) <sub>4</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH(Ph)	0.45	0.11
50	NH(CH <sub>2</sub> ) <sub>3</sub> NMe <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	0.68	0.075
51	NMe(CH <sub>2</sub> ) <sub>3</sub> NMe <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	16	2.0
52	NHCH <sub>2</sub> CMe <sub>2</sub> CH <sub>2</sub> NMe <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	3.2	0.21
53	NH(CH <sub>2</sub> ) <sub>3</sub> (morpholin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	0.84	0.072
54	NH(CH <sub>2</sub> ) <sub>3</sub> (2-metilpiperidin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	0.73	0.060
55	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NH <sub>2</sub>	9.6	0.45
56	NH(CH <sub>2</sub> ) <sub>4</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	0.28	0.035
57	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH <sub>2</sub>	2.4	0.14
58	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(Et)	0.42	0.053
59	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(alil)	0.76	0.035
60	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>i</i> -Pr)	0.55	0.034
61	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>n</i> -octil)	6.2	0.42
62	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(bencil)	2.5	0.062
63	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(ciclohexil)	0.37	0.029
64	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(BOCNH(CH <sub>2</sub> ) <sub>2</sub> )	4.2	0.067
65	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(Me <sub>2</sub> N(CH <sub>2</sub> ) <sub>2</sub> )	2.4	0.075
66	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCON(Et <sub>2</sub> )	>50	5.5
67	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONHPh	0.57	0.084
68	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(4-ClPh)	1.8	0.12
69	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(4-BrPh)	1.5	0.11
70	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(4-CF <sub>3</sub> Ph)	5.5	0.51
71	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(3,4-(Cl) <sub>2</sub> Ph)	5.0	0.60
72	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(4-MePh)	0.84	0.11
73	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(2-MeOPh)	0.66	0.067
74	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(3-MeOPh)	0.91	0.063
75	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(4-MeOPh)	0.68	0.074
76	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH(1-naftil)	2.6	0.30
77	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCOCH <sub>2</sub> ( <i>t</i> -Bu)	>50 <sup>b</sup>	27
78	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCHNMe <sub>2</sub>	>50	10
79	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCSNHEt	5.0	0.26
80	NH(CH <sub>2</sub> ) <sub>3</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHNHCNHEt	>50	2.7
81	NH(CH <sub>2</sub> ) <sub>4</sub> NEt <sub>2</sub>	2,6-(Cl) <sub>2</sub> Ph	NHCSNH(CH <sub>2</sub> ) <sub>3</sub> morfolin-1-il	1.1	0.13
82	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCOCH <sub>2</sub> Et	>50	3.7
83	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCOCH <sub>2</sub> ( <i>t</i> -Bu)	>50	6.8
84	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCONH( <i>t</i> -Bu)	0.47	0.051
85	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCOCH <sub>2</sub> (Ph)	>50	6.8
86	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHSO <sub>2</sub> (Ph)	6.9	0.068
87	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHNHCNHPH	37	2.8
88	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCN( <i>i</i> -Pr)NH( <i>i</i> -Pr)	>50	2.5
89	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCSNH( <i>t</i> -Bu)	1.9	0.10
90	NH(CH <sub>2</sub> ) <sub>3</sub> (N-metilpiperazin-1-il)	2,6-(Cl) <sub>2</sub> Ph	NHCSNHPh	2.2	0.32

## Indolin-2-onas:

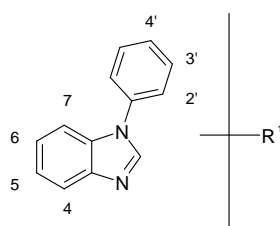


Nº	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	IC <sub>50</sub> (µM) PDGFR	IC <sub>50</sub> (µM) FGFR
1	H	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	H	0.51	0.03
2	4-CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	H	2.11	0.03
3	5-Br	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	H	0.62	0.08
4	6-(3-OCH <sub>3</sub> fenil)	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	H	0.45	1.2
5	6-(3-OC <sub>2</sub> H <sub>5</sub> fenil)	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	H	2.38	4.5
6	H	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	H	0.14	3.68
7	5-COOH	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	H	4.19	0.77
8	5-SO <sub>2</sub> -NH <sub>2</sub>	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	H	5.98	0.46
9	6-OCH <sub>3</sub>	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	H	0.14	3.89
10	6-fenil	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	H	0.16	1.05
11	6-(3-OCH <sub>3</sub> fenil)	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	H	0.17	0.36
12	6-(2-OCH <sub>3</sub> fenil)	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	H	1.42	2.13
13	6-(4-OCH <sub>3</sub> fenil)	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	H	3.08	1.54
14	H	CH <sub>3</sub>	H	CH <sub>3</sub>	10.50	7.08
15	H	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	0.06	3.04
16	5-Br	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	0.06	2.05
17	5-COOH	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	1.21	0.28
18	5-SO <sub>2</sub> -NH <sub>2</sub>	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	1.53	0.28
19	6-OCH <sub>3</sub>	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	0.55	5.40
20	6-fenil	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	0.01	2.29
21	6-(3-OCH <sub>3</sub> fenil)	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	0.10	1.40
22	6-(2-OCH <sub>3</sub> fenil)	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	2.22	7.33
23	6-(4-OCH <sub>3</sub> fenil)	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	CH <sub>3</sub>	1.0	5.76



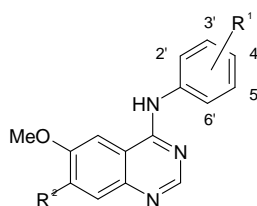
Nº	R <sup>1</sup>	R <sup>2</sup>	IC <sub>50</sub> (µM) PDGFR	IC <sub>50</sub> (µM) FGFR	IC <sub>50</sub> (µM) EGFR
24	H	H	0.24	0.27	>100
25	5-Br	H	0.004	0.27	>100
26	5-COOH	H	<0.78	0.20	29.1
27	5-SO <sub>2</sub> -NH <sub>2</sub>	H	0.02	0.22	56.5
28	6-OCH <sub>3</sub>	H	19.3	1.08	>100
29	6-fenil	H	0.04	1.27	>100
30	6-(3-OCH <sub>3</sub> fenil)	H	0.68	1.35	8.88
31	6-(2-OCH <sub>3</sub> fenil)	H	0.87	0.08	>100
32	6-(4-OCH <sub>3</sub> fenil)	H	5.43	0.88	>100
33	H	CH <sub>2</sub> CH <sub>2</sub> COOH	5.46	10.5	>100
34	5-Br	CH <sub>2</sub> CH <sub>2</sub> COOH	0.92	13.3	>100
35	5-COOH	CH <sub>2</sub> CH <sub>2</sub> COOH	11.5	0.22	>100
36	5-SO <sub>2</sub> -NH <sub>2</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	1.03	0.31	>100
37	6-OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> COOH	3.46	20	>100
38	6-fenil	CH <sub>2</sub> CH <sub>2</sub> COOH	70.6	20	>100
39	6-(3-OCH <sub>3</sub> fenil)	CH <sub>2</sub> CH <sub>2</sub> COOH	>100	17.5	>100
40	6-(2-OCH <sub>3</sub> fenil)	CH <sub>2</sub> CH <sub>2</sub> COOH	13.9	17.7	>100
41	6-(4-OCH <sub>3</sub> fenil)	CH <sub>2</sub> CH <sub>2</sub> COOH	94.01	20	>100

## 1-fenilbenzimidazoles:



Nº	R <sup>1</sup>	IC <sub>50</sub>	IC <sub>50</sub>
		(µM) PDGFR	(µM) FGFR
1	5-OH	0.44	6.4
2	5-OMe	0.43	22
3	5-OEt	0.24	26
4	5-O- <i>n</i> Pr	0.25	35
5	5-O- <i>i</i> Pr	3.1	50
6	5-O- <i>n</i> Bu	1.3	50
7	5-OCH <sub>2</sub> CH=CH <sub>2</sub>	0.61	35
8	5-O(CH <sub>2</sub> ) <sub>4</sub> OH	0.45	25
9	5-OCH <sub>2</sub> (oxiranil)	0.32	18
10	5-OCH <sub>2</sub> CH(OH)CH <sub>2</sub> OH	0.31	15
11	5-O(CH <sub>2</sub> ) <sub>2</sub> NH <sub>2</sub>	0.65	27
12	5-O(CH <sub>2</sub> ) <sub>2</sub> NMe <sub>2</sub>	1.5	45
13	5-O(CH <sub>2</sub> ) <sub>3</sub> NMe <sub>2</sub>	0.15	4.1
14	5-O(CH <sub>2</sub> ) <sub>4</sub> NMe <sub>2</sub>	0.16	6.8
15	5-O(CH <sub>2</sub> ) <sub>2</sub> Nmorph	0.73	28
16	5-O(CH <sub>2</sub> ) <sub>3</sub> Nmorph	0.17	9.8
17	5-O(CH <sub>2</sub> ) <sub>4</sub> Nmorph	0.27	12
18	5-OMe, 4'-NH <sub>2</sub>	0.28	18
19	5-SMe	0.74	27
20	5-S(CH <sub>2</sub> ) <sub>3</sub> Nmorph	>50	24
21	4,5-diOH	25	1.9
22	5,6-diOMe	1.2	25
23	5,6-diOH	2.3	11
24	5,6-OCH <sub>2</sub> O	2.2	27
25	5-OH, 6-COOH	4.3	22
26	5-OMe, 6-COOMe	0.87	50
27	5-OMe, 6-CH <sub>2</sub> OH	0.37	11
28	5-OMe, 6-CHO	1.0	25

## 4-anilinoquinazolininas:



N°	R <sup>1</sup>				R <sup>2</sup>	IC <sub>50</sub> (μM) FGFR	IC <sub>50</sub> (μM) EFGR
	2'	4'	5'	6'			
1	F	Br	H	H	1-(1,2,3-triazolil)-(CH <sub>2</sub> ) <sub>2</sub> O	>100	0.4
2	F	Br	H	H	MeN(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> CH-CH <sub>2</sub> O	3.0	0.5
3	F	Cl	H	H	MeO(CH <sub>2</sub> ) <sub>2</sub> O	27	0.1
4	F	Cl	H	H	4-Me-piperazinil-(CH <sub>2</sub> ) <sub>3</sub> O	2.1	0.075
5	F	Br	H	H	4-Me-piperazinil-(CH <sub>2</sub> ) <sub>3</sub> O	2.0	0.1
6	F	Cl	H	H	4-Me-piperazinil-(CH <sub>2</sub> ) <sub>2</sub> O	8	0.2
7	F	Cl	H	H	4-morfolinil-(CH <sub>2</sub> ) <sub>3</sub> O	6.7	0.1
8	F	Cl	H	H	4-morfolinil-(CH <sub>2</sub> ) <sub>2</sub> O	19	0.4
9	F	Cl	H	H	1-pirrolidinil-(CH <sub>2</sub> ) <sub>3</sub> O	4.2	0.1
10	F	Cl	H	H	(CH <sub>2</sub> ) <sub>4</sub> N-CH <sub>2</sub> CH=CH-CH <sub>2</sub> O	1.9	0.09
11	F	Br	H	H	(CH <sub>2</sub> ) <sub>4</sub> N-CH <sub>2</sub> CH=CH-CH <sub>2</sub> O	2.1	0.1
12	F	CN	H	H	(CH <sub>2</sub> ) <sub>4</sub> N-CH <sub>2</sub> CH=CH-CH <sub>2</sub> O	19	0.8
13	F	Cl	H	H	4-piridil-N(Me)-(CH <sub>2</sub> ) <sub>2</sub> O	1.6	0.3
14	F	Cl	H	H	MeN(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> CH-O	50	nt
15	F	Cl	H	H	MeN(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> CH-CH <sub>2</sub> O	4.3	0.3
16	F	Me	H	H	MeN(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> CH-CH <sub>2</sub> O	13.5	1
17	F	Cl	OH	H	MeN(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> CH-CH <sub>2</sub> O	0.9	0.1
18	F	Me	OH	H	MeN(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> CH-CH <sub>2</sub> O	1.5	0.3
19	F	Cl	H	F	MeN(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> CH-CH <sub>2</sub> O	1.7	0.45
20	F	Br	H	F	MeN(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> CH-CH <sub>2</sub> O	1.5	0.3
21	F	Cl	H	H	HN(CH <sub>2</sub> CH <sub>2</sub> )CH-CH <sub>2</sub> O	5.0	0.3
22	F	Br	H	H	HN(CH <sub>2</sub> CH <sub>2</sub> )CH-CH <sub>2</sub> O	6	0.2
23	F	Me	H	H	HN(CH <sub>2</sub> CH <sub>2</sub> )CH-CH <sub>2</sub> O	12.3	0.5
24	F	Cl	H	F	HN(CH <sub>2</sub> CH <sub>2</sub> )CH-CH <sub>2</sub> O	2.5	0.5
25	F	Br	H	F	HN(CH <sub>2</sub> CH <sub>2</sub> )CH-CH <sub>2</sub> O	1.9	0.45
26	F	Cl	H	H	MeN(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> CH-CH <sub>2</sub> CH <sub>2</sub> O	4.1	0.3
27	F	Br	H	H	MeN(CH <sub>2</sub> CH <sub>2</sub> ) <sub>2</sub> CH-CH <sub>2</sub> CH <sub>2</sub> O	2.1	0.2
28	F	Br	H	H	HN(CH <sub>2</sub> CH <sub>2</sub> )CH-CH <sub>2</sub> CH <sub>2</sub> O	2.7	0.2
29	F	Cl	H	H	(R) MeN(CH <sub>2</sub> )(CH <sub>2</sub> ) <sub>3</sub> CH-CH <sub>2</sub> O	1.8	0.1
30	F	Br	H	H	(R) MeN(CH <sub>2</sub> )(CH <sub>2</sub> ) <sub>3</sub> CH-CH <sub>2</sub> O	1.8	0.2



---

**Bibliografía**

- 1 Alvarez, J.; Shoichet, B. En *Virtual Screening in Drug Discovery*; Alvarez, J.; Shoichet, B., Ed.; Taylor & Francis Group. Florida, 2005.
- 2 Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science*. **2004**, *303*(19), 1813-1818.
- 3 Böhm, H-J.; Schneider, G.; Kubinyi, H.; Mannhold, R.; Timmerman, H. En *Virtual Screening for Bioactive Molecules* (Methods and Principles in Medicinal Chemistry). Ed.; Wiley-VCH: Weinheim, 2000; Vol. 10.
- 4 Venter, J. C.; Adams, M. D.; Myers, E. W. ; Li, P. W.; Mural, R. J.; Sutton, G. G.; *et al.* The sequence of the human genome. *Science*. **2001**, *291*, 1304-1351.
- 5 Böhm, H. J.; Boehringer, M.; Bur, D.; Gmuender, H.; Huber, W.; Klaus, W.; Kostrewa, D.; Kuehne, H.; Luebbers, T.; Meunier-Keller, N.; Mueller, F. Novel Inhibitors of DNA Gyrase: 3D Structure Based Biased Needle Screening, Hit Validation by Biophysical Methods, and 3D Guided Optimization. A Promising Alternative to Random Screening. *J. Med. Chem.* **2000**, *43*, 2664-2674
- 6 Schmid, E.; Smith, D. A. Keynote review: Is declining innovation in the pharmaceutical industry a myth?. *Drug Discovery Today*. **2005**, *10*(15), 1031-1039.
- 7 Oprea, T. I.; En *Cheminformatics-from Data to Knowledge*; Gasteiger, J.; Engels, T., Ed. Wiley-VCH: New York, 2003; Vol 2, pp 1508-1531.
- 8 Leo, A.; Hansch, C.; Church, C. Comparison of parameters currently used in the study of structure-activity relationships. *J. Med. Chem.* **1969**, *12*, 766-771.
- 9 Terrett, N. K.; Bell, S. A.; Brown, D. Sildenafil, P. E. (Viagra), a potent and selective inhibitor of Type 5cGMP phosphodiesterase with utility for the treatment of male erectile dysfunction. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1819-1824.
- 10 Terfloth, L. *Drug Design*. En *Cheminformatics*; Gasteiger, J.; Engels, T., Ed.; Wiley-VCH: Weinheim, 2003. pp 597-618.
- 11 Comprehensive Medicinal Chemistry (CMC), Molecular Design Limited, San Leandro, CA, <http://www.mdli.com>.
- 12 MACCS-II Drug Data Report (MDDR), Molecular Design Limited, San Leandro, CA, <http://www.mdli.com>.
- 13 World Drug Index (WDI), Derwent Information, London, UK, <http://www.derwent.com>.
- 14 Available Chemicals Directory (ACD). MDL Information Systems Inc., San Leandro, CA, <http://www.mdli.com>.
- 15 Walters, W.; Murcko, M. A. Prediction of 'drug-likeness'. *Adv. Drug Deliv. Rev.* **2002**, *54*, 255-271.
- 16 Lipinski, C. A.; Lombardo, F.; Dominy, B.W.; Feeny, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3-25.
- 17 Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput Aided Mol. Des.* **2000**, *14*, 251-264.

- 18 Muegge, I.; Heald, S. L.; Brittelli, D. Simple selection criteria properties for drug-like chemical matter. *J. Med. Chem.* **2001**, *44*, 1841–1846.
- 19 Ajay, W.; Walters, P.; Murcko, M. A. Can we learn to distinguish between ‘drug-like’ and ‘nondrug-like’ molecules. *J. Med. Chem.* **1998**, *41(18)*, 3314–3324.
- 20 Sadowski, J.; Kubinyi, H. A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* **1998**, *41(18)*, 3325–3329.
- 21 Gillet, V. J.; Willett, P.; Bradshaw, J. Identification of Biological Activity Profiles Using Substructural Analysis and Genetic algorithms. *J. Chem. Inf. Comput. Sci.* **1998**, *38(2)*, 165-179.
- 22 Oprea, T. I.; Matter, H. Integrating virtual screening in lead discovery. *Curr. Opin. Chem. Biol.* **2004**, *8(4)*, 349-358.
- 23 Johnson, M. A.; Maggiora, G. M. En *Concepts and Application of Molecular Similarity*. Ed.; Wiley: New York, 1990.
- 24 Martin, Y. C.; Kofron J. L.; Traphagen L. M. Do structurally similar molecules have similar biological activity?. *J. Med. Chem.* **2002**, *45(19)*, 4350-4358.
- 25 Martin, Y. C. Diverse viewpoints on computational aspects of molecular diversity. *J. Comb. Chem.* **2001**, *3(3)*, 231-50.
- 26 Willet, P. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38(6)*, 983-996.
- 27 Downs, G. M.; Willett, P. Similarity Searching in Databases of Chemical Structures. En *Reviews in Computational Chemistry*, Lipkowitz, K. B.; Boyd, D. B., Ed.; Wiley and Sons: New York, 1995; Vol 7, pp 1-66.
- 28 Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Ed.; Wiley-VCH: Weinheim, 2000.
- 29 Brown, R. D.; Martin, Y. C. Use of Structure-Activity Data To Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572-584.
- 30 Maccs-II, Molecular Design Limited, San Leandro, CA.
- 31 Daylight Chemical Information Software, Ver 4.41; Daylight Chemical Information, Inc., Irvine, CA.
- 32 Unity Chemical Information Software Ver 2.3; Tripos Associates, St. Louis.
- 33 Broto, P.; Moreau, G.; Vandyke, C. Molecular structures: Perception, autocorrelation descriptor and SAR studies. *Eur. J. Med. Chem.* **1984**, *19*, 66-70.
- 34 Schneider, G.; Renner, S.; Fechner, U. *Navigation in Chemical Space Based On Correlation-Vector Representation of Molecules*. En Proceedings of the Beilstein-Workshop 2004; Kettner, K.; Hicks, M.; Ed.; Logos Verlag: Berlin, 2004.
- 35 Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRIND-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43(17)*, 233-43.
- 36 Bauknecht, H.; Zell, A.; Bayer, H.; Levi, P.; Wagener, M.; Sadowski, J.; Gasteiger, J. Locating biologically active compounds in medium-sized heterogeneous datasets by topological autocorrelation vectors: dopamine and benzodiazepine agonists. *J. Chem. Inf. Comput. Sci.* **1996**, *36(6)*, 1205-1213.



- 37 Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood behavior: a useful concept for validation of "molecular diversity" descriptors. *J. Med. Chem.* **1996**, *39*(16), 3049-3059.
- 38 MOE; Chemical Computing Group, Inc. 1010 Serbrooke Street West, Suite 910.
- 39 Accelrys, Inc. 9685 Scranton Road Sand Diego, CA 92121-3752.
- 40 Milano Chemometrics and QSAR Research Group. University of Milano-Bicocca, Department of Environmental Sciences. Piazza della Scienza, 1-20126. Milano.
- 41 Norgwyn Montgomery Software Inc. 216 Lower Valley Rd. North Wales, PA 19454.
- 42 CambridgeSoft Corp. 100 Cambridge, MA 02140.
- 43 Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold-Hopping" by topological pharmacophore search: a contribution to virtual screening. *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 2894-2896.
- 44 Schneider, G.; Nettekoven, M. Ligand-based combinatorial design of selective purinergic receptor (A2A) antagonists using self-organizing maps. *J. Comb. Chem.* **2003**, *5*(3), 223-237.
- 45 Naerum, L.; Norskov-Lauritsen, L.; Olesen, P. H. Scaffold hopping and optimization towards libraries of glycogen synthase kinase-3 inhibitors. *Bioorg. Med. Chem. Lett.* **2002**, *12*(11), 1525-1528.
- 46 Gund, P. En *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: La Jolla, 2000, Chapter 1, pp 3-11.
- 47 Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A fast approach to pharmacophore mapping against application to dopaminergic and benzodiazepine agonists. *J. Comput Aided Mol. Des.* **1993**, *7*, 83-102.
- 48 Barnum, D.; Greene, J.; Smellie, A.; Sprage, P. Identification of common functional configurations among molecules. *J. Chem. Inf. Comput. Sci.* **1996**, *36*(3), 563-571.
- 49 Jones, G.; Willet, P.; Glen, R. C. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput Aided Mol. Des.* **1995**, *9*(6), 532-549.
- 50 Van Drie, J. H.; Weininger, D.; Martin, Y. C. ALADDIN: an integrated tool for computer-assisted molecular design and pharmacophore recognition from geometric, steric, and substructure searching of three-dimensional molecular structures. *J. Comput Aided Mol. Des.* **1989**, *3*(3), 225-251.
- 51 Renner, S.; Schneider, G. Fuzzy Pharmacophore Models from Molecular Alignments for Correlation-Vector-Based Virtual Screening. *J. Med. Chem.* **2004**, *47*, 4653-4664.
- 52 Renner, S.; Ludwig, V.; Boden, O.; Scheffer, U.; Göbel, M.; Schneider, G. New Inhibitors of the Tat-TAR RNA Interaction Found with a "Fuzzy" Pharmacophore Model. *ChemBioChem.* **2005**, *6*(6), 1-8.
- 53 Marriott, D. P.; Dougall, I. G.; Meghani, P.; Liu, Y. J.; Flower, D. R. Lead generation using pharmacophore mapping and three-dimensional database searching: application to muscarinic M(3) receptor antagonists. *J. Med. Chem.* **1999**, *42*(17), 3210-3206.
- 54 Singh, J.; Van Vlijmen, H.; Liao, Y.; *et al.* Identification of Potent and Novel alpha4beta1 Antagonists Using in Silico Screening. *J. Med. Chem.* **2002**, *45*(14), 2988-2993.

- 55 Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Amer. Chem. Soc.* **1998**, *110*(18), 5959-5967.
- 56 Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*(7), 849-857.
- 57 Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant. Struc.–Act. Relat.* **1993**, *12*, 9-20.
- 58 Beddell, C. R.; Goodford, P. J.; Norrington, F. E.; Wilkinson, S.; Wootton, R. Compounds designed to fit a site of known structure in human haemoglobin. *Br. J. Pharmacol.* **1976**, *57*, 201–209.
- 59 Cohen, S. S. A strategy for the chemotherapy of infectious disease. *Science.* **1977**, *197*, 431–432.
- 60 Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer E. F. Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*(3), 535-542.
- 61 Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science.* **1992**, *257*, 1078–1082.
- 62 Brenk, R.; Naerum, L.; Gradler, U.; Gerber, H. D.; Garcia, G. A.; Reuter, K.; Stubbs, M. T.; Klebe, G. Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J. Med. Chem.* **2003**, *46*, 1133-1143.
- 63 Gruneberg, S.; Stubbs, M. T.; Klebe, G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J. Med. Chem.* **2002**, *45*(17), 3588-3602.
- 64 Kraemer, O.; Hazemann, I.; Podjarny, A. D.; Klebe, G. Virtual screening for inhibitors of human aldose reductase. *Proteins.* **2004**, *55*(4), 814-823.
- 65 Powers, R. A.; Morandi, F.; Shoichet, B. K. Structure-based discovery of a novel, noncovalent inhibitor of AmpC betalactamase. *Structure.* **2002**, *10*(7), 1013-1023.
- 66 Soelaiman, S.; Wei, B. Q.; Bergson, P.; Lee, Y. S.; Shen, Y.; Mrksich, M.; Shoichet, B. K.; Tang, W. J. Structure-based inhibitor discovery against adenylyl cyclase toxins from pathogenic bacteria that cause anthrax and whooping cough. *J. Biol. Chem.* **2003**, *278*(28), 25990-25997.
- 67 Liu, H.; Li, Y.; Song, M.; Tan, X.; Cheng, F.; Zheng, S.; Shen, J.; Luo, X.; Ji, R.; Yue, J.; Hu, G.; Jiang, H.; Chen, K. Structure based discovery of potassium channel blockers from natural products: virtual screening and electrophysiological assay testing. *Chem. Biol.* **2003**, *10*(11), 1103-1113.
- 68 Vangrevelinghe, E.; Zimmermann, K.; Schoepfer, J.; Portmann, R.; Fabbro, D.; Furet, P. Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *J. Med. Chem.* **2003**, *46*(13), 2656-2662.
- 69 Evers, A.; Klebe, G. Ligand-supported homology modeling of G-protein-coupled receptor sites: models sufficient for successful virtual screening. *Angew. Chem. Int. Ed. Engl.* **2004**, *43*(2), 248–251.
- 70 Hetényi, C.; Van der Spoel, D. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Science.* **2002**, *11*, 1729-1737.

- 71 Muegge, I.; Rarey, M. Small Molecule Docking and Scoring. En *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Ed.; Wiley and Sons: New York, 2001; Vol 17, pp 1-60.
- 72 Bissantz, C.; Folker, G.; Rognan, D. Protein-based virtual screening of chemical databases 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*(25), 4759-4767.
- 73 Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*(2), 269-288.
- 74 Perola, E.; Xu, K.; Kollmeyer, T. M.; Kaufmann, S. H.; Prendergast, F. G.; Pang, Y. P. Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J. Med. Chem.* **2000**, *43*(3), 401-408.
- 75 Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.* **2003**, *21*(4), 289-307.
- 76 Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*(3), 470-89.
- 77 Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **1996**, *3*(6), 449-462.
- 78 Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins.* **1998**, *33*, 367-382.
- 79 Hou, T.; Wang, J.; Chen, L.; Xu, X. Automated docking of peptides and proteins by using a genetic algorithm combined with a tabu search. *Protein. Eng.* **1999**, *12*(8), 639-647.
- 80 Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*(3), 727-748.
- 81 Morris, G. M.; Goodsell, D. S.; Halliday, R.; Huey, R.; Hart, W.E.; Belew, R. K.; Olson, A. J. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*(14), 1639-1662.
- 82 Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*(25), 5100-5109.
- 83 Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2*, 317-324.
- 84 Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins.* **1990**, *8*, 195-202.
- 85 Glide, Schrödinger Inc., New York, 2003.
- 86 Liu, M.; Wang, S. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. *J. Comput Aided Mol. Des.* **1999**, *13*(5), 435-451.
- 87 McMartin, C.; Bohacek, R. S. QXP: Powerful, rapid computer algorithms for structure-based drug design. *J. Comput Aided Mol. Des.* **1997**, *11*(4), 333-344.
- 88 Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM- A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994**, *15*, 488-506.
- 89 Metaphorics LLC, Piemont, CA 94611.

- 90 Böhm, H. J.; Stahl, M. The use of scoring functions in drug discovery applications. En *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Ed.; Wiley and Sons: New York, 2002; Vol 18, pp 41-87.
- 91 Momany, F. A.; Rone, R. Validation of the general-purpose QUANTA. 3.2/CHARMm force-field. *J. Comput. Chem.* **1992**, *13*(7), 888-900.
- 92 Böhm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput Aided Mol. Des.* **1994**, *8*, 243-256.
- 93 Eldridge, M.; Murray, C. W.; Auton, T. A.; Paolini, G. V.; Lee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput Aided Mol. Des.* **1997**, *11*(5), 425-445.
- 94 Wang, R.; Liu, L.; Lai, L.; Tang, Y. SCORE: a new empirical method for estimating the binding affinity of a protein-ligand complex. *J. Mol. Model.* **1998**, *4*, 379-394.
- 95 Rognan, D.; Laumoeller, S. L.; Holm, A.; Buus, S.; Tschinke, V. Predicting binding affinities of protein ligands from threedimensional coordinates: Application to peptide binding to class I major histocompatibility proteins. *J. Med. Chem.* **1999**, *42*, 4650-4658.
- 96 Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*(5), 791-804.
- 97 Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*(2), 337-356.
- 98 DeWitte, R. S.; Shakhnovich, E. I. SmoG: de Novo Design Method based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *J. Am. Chem. Soc.* **1996**, *118*, 11733-11744.
- 99 Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Thornton, J. M. BLEEP –Potential of Mean Force Describing Protein-Ligand Interactions: I. Generating Potential. *J. Comput. Chem.* **1999**, *20*(11), 1165-1176.
- 100 Ishchenko, A. V.; Shakhnovich, E. I. Small Molecule Growth 2001(SMoG2001): An Improved Knowledge-based Scoring Function for Protein-Ligand Interactions. *J. Am. Chem. Soc.* **2002**, *45*(13), 2770-2780.
- 101 Shoichet, B. K. Virtual Screening of Chemical Libraries. *Nature.* **2004**, *432*(7019), 40-43.
- 102 Mestres, J.; Knegt, R. Similarity versus docking in 3D virtual screening. *Persp. Drug. Disc. Design.* **2000**, *20*(1), 191-207.
- 103 Claussen, H.; Bunning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*(2), 377-395.
- 104 Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: incorporation of protein mobility an structural water heterogeneity in AutoDock. *Proteins.* **2002**, *46*, 34-40.
- 105 Dixon, J. S. Evaluation of the CASP2 docking section. *Proteins* **1997**, *Suppl*, 198-204.
- 106 Schneider, G.; Böhm, H.-J. Virtual screening and fast automated docking methods. *DDT.* **2002**, *7*(1), 64-70.
- 107 Böhm, H. J.; Current computational tools for *de novo* ligand design. *Curr. Opin. Biotechnol.* **1996**, *7*, 433-436.

- 108 Schneider, G.; So, S-S. Evolutionary *de novo* design. En Adaptive Systems in Drug Design; Schneider, G., Ed.; Landes Biosciences: 2002.
- 109 Böhm, H. J.; The computer program LUDI: a new method for the *de novo* design of enzyme inhibitors. *J. Comput Aided Mol. Des.* **1992**, *6*(1), 61-78.
- 110 Roe, D. C.; Kuntz, I. D. BUILDER v.2: Improving the chemistry of a *de novo* design strategy. *J. Comput Aided Mol. Des.* **1995**, *9*(3), 269-282.
- 111 Lauri, G.; Bartlett, P. A. CAVEAT: A program to facilitate the design of organic molecules. *J. Comput Aided Mol. Des.* **1994**, *8*(1), 51-66.
- 112 Gillet, V.; Johnson, A. P.; Mata, P.; Sike, S.; Williams, P. SPROUT: a program for structure generation. *J. Comput Aided Mol. Des.* **1993**, *7*(2), 127-153.
- 113 Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. Thermodynamics of cavity formation in water. *Faraday Symp. Chem. Soc.* **1982**, *17*, 55-67.
- 114 Tembe, B. L.; McCammon, J. A. Ligand-receptor interactions. *Comput. Chem.* **1984**, *8*, 281-283.
- 115 Jorgensen, W. L.; Ravimohan, C. Monte Carlo simulation of differences in free energies of hydration. *J. Chem. Phys.* **1985**, *83*, 3050-3054.
- 116 Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. 3rd. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*(12), 889-97.
- 117 Aqvist, J.; Medina, C.; Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*(3), 385-391.
- 118 Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401-9409.
- 119 Kuhn, B.; Kollman, P. A.; A Ligand That Is Predicted to Bind Better to Avidin than Biotin: Insights from Computational Fluorine Scanning. *J. Am. Chem. Soc.* **2000**, *122*(16), 3909-3916.
- 120 Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Wang, J.; Luo, R.; Bayly, C. I.; Kollman, P. A. An Analysis of the Interactions between the Sem-5 SH3 Domain and Its Ligands Using Molecular Dynamics, Free Energy Calculations, and Sequence Analysis. *J Am Chem Soc.* **2001**, *123*(17), 3986-94.
- 121 Pearlman, D. A. Evaluating the Molecular Mechanics Poisson-Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to p38 MAP Kinase. *J. Med. Chem.* **2005**, *48*(24), 7796-7807.
- 122 Kuhn, B.; Kollman, P. A. Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* **2000**, *43*(20), 3786-3791.
- 123 Donini, O. A. T.; Kollman, P. A. Calculation and Prediction of Binding Free Energies for the Matrix Metalloproteinases. *J. Med. Chem.* **2000**, *43*(22), 4180-4188.
- 124 Zhuang, S.; Zou, J.; Jiang, Y.; Mao, X.; Zhang, B.; Liu, H.; Yu, Q. Some Insights into the Stereochemistry of Inhibition of Macrophage Migration Inhibitory Factor with 2-Fluoro-*p*-hydroxycinnamate and Its Analogues from Molecular Dynamics Simulations. *J. Med. Chem.* **2005**, *48*(23), 7208-7214.

- 125 Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and Use of the MM-PBSA Approach for Drug Discovery. *J. Med. Chem.* **2005**, *48*(12), 4040-4048.
- 126 *Proteins Structure Prediction. Methods and Protocols.* Methods in Molecular Biology; Webster, D.M., Ed.; Humana Press: UK: 2000, pp 1-417.
- 127 Chothia, C. Proteins. One thousand families for the molecular biologist. *Nature.* **1992**, *357*(6379), 543-544.
- 128 Holm, L.; Sander, C. Mapping the Protein Universe. *Science.* **1996**, *273*, 595-602.
- 129 Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.* **1999**, *12*, 85-94.
- 130 Sánchez, R.; Sali, A. Advances in Comparative protein-structure modelling. *Curr. Opin. Struct. Biol.* **1997**, *7*, 206-214.
- 131 Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779-815.
- 132 Insight II. Accelrys, Inc. 9685 Scranton Road Sand Diego, CA 92121-3752. Accesible a través de la dirección <http://www.accelrys.com/products/insight/>. En línea (12/1/06).
- 133 Swiss-Model. Accesible a través de la dirección <http://swissmodel.expasy.org//SWISS-MODEL.html>. En línea (12/1/06).
- 134 CPHmodels 2.0 Server. Accesible a través de la dirección <http://www.cbs.dtu.dk/services/CPHmodels/>. En línea (12/1/06)
- 135 WHAT IF Web Interface. Accesible a través de la dirección <http://swift.cmbi.kun.nl/WIWWWI/>. En línea (12/1/06).
- 136 Fiser, A.; Sali, A. Modeling of loops in protein structures. *Protein Science.* **2000**, *9*, 1753-1773.
- 137 Simons, K.; Kooperber, Ch.; Huang, E.; Baker, D. Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences using Simulated Annealing and Bayesian Scoring Functions. *J. Mol. Biol.* **1997**, *268*, 209-225. Accesible a través de la dirección: <http://www.bioinfo.rpi.edu/~bystrc/hmmstr/server.php>. En línea (12/1/06).
- 138 Jones, D. T.; Taylor, W. R.; Thornton, J. M. A New Approach to Protein Recognition. *Nature.* **1992**, *358*, 86-89.
- 139 Ho, C. M.; Marshall, G. R. Cavity search: an algorithm for the isolation and display of cavity-like binding regions. *J. Comput Aided Mol. Des.* **1990**, *4*(4), 337-354.
- 140 Lim-Wilby, M.; Lyons, T.; Dooley, M.; Goupil-Lamy, A.; Patel, S.; Schneider, C.; Hoffmann, R.; Bertrand, O.; Güner, O.; Classification of Ligand-Receptor Complexes Based on Receptor Binding Site Characteristics. En *Virtual Screening in Drug Discovery*; Alvarez, J.; Shoichet, B., Ed.; Taylor & Francis Group: Florida, 2005, pp 279-299.
- 141 Fradera, X.; Mestres, J. Guided Docking Approaches to Structure-Based Design and Screening. *Curr. Top. Med. Chem.* **2004**, *4*(7), 687-700.
- 142 Evers, A.; Hessler, G.; Matter, H.; Klabunde, T. Virtual Screening of Biogenic Amine-Binding G-Protein Coupled Receptors: Comparative Evaluation of Protein- and Ligand-Based Virtual Screening Protocols. *J. Med. Chem.* **2005**, *48*, 5448-5465.
- 143 Venkatachalan, C. M.; Kirchhoff, P.; Waldman, M. Receptor-Based Pharmacophore Perception and Modeling. En *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: La Jolla, 2000, Chapter 18, pp 341-350.

- 144 Wang, S.; Milne, G. W.; Yan, X.; Posey, I. J.; Nicklaus, M. C.; Graham, L.; Rice, W. G. Discovery of novel, non-peptide HIV-1 protease inhibitors by pharmacophore searching. *J. Med. Chem.* **1996**, *39*(10), 2047-54.
- 145 Deng, Z.; Chuaqui, C.; Singh, J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* **2004**, *47*(2), 337-44.
- 146 Valler, M. J.; Green, D. Diversity screening versus focussed screening in drug discovery. *DDT*. **2000**, *5*(7), 286-293.
- 147 Willet, P. Chemoinformatics – similarity and diversity in chemical libraries. *Curr. Opin. Biotechnol.* **2000**, *11*, 85–88.
- 148 Perez, J. J. Managing molecular diversity. *Chem. Soc. Rev.* **2005**, *34*, 143–152.
- 149 Weber, L. Current Status of Virtual Combinatorial Library Design. *QSAR & Comb. Sci.* **2005**, *24*, 809-823.
- 150 Gillet, V. J.; Willett, P.; Bradshaw, J. The Effectiveness of Reactant Pools for Generating Structurally-Diverse Combinatorial Libraries. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 731-740.
- 151 Gillet, V. J.; Nicolotti, O. Evaluation of reactant-based and product-based approaches to the design of combinatorial libraries. *Perspect Drug Discov.* **2000**, *20*, 265-287.
- 152 Jamois, E. A.; Hassan, M.; Waldman, M. Evaluation of Reagent-Based and Product-Based Strategies in the Design of Combinatorial Library Subsets. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 63-70.
- 153 Gillet, V. J.; Willett, P.; Bradshaw, J.; Green V. S. D. Selecting Combinatorial Libraries to Optimize Diversity and Physical Properties. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 169-177.
- 154 Brown, R.; Martin, Y. C. Designing Combinatorial Library Mixtures Using a Genetic Algorithm. *J. Med. Chem.* **1997**, *40*, 2304-2313.
- 155 Good, A.C.; Lewi, R.A. New Methodology for Profiling Combinatorial Libraries and Screening Sets: Cleaning Up de Design Process with HARPick. *J. Med. Chem.* **1997**, *40*, 3926-3936.
- 156 Sheridan, R. P.; Kearsley, S. K. Using a Genetic Algorithm to Suggest Combinatorial Libraries. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 310-320.
- 157 Weber, L.; Wallbaum, S.; Broger, C.; Gubernator, K. Optimization of the Biological Activity of Combinatorial Compound Libraries by a Genetic Algorithm. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*(20), 2280-2282.
- 158 Sheridan, R. P.; SanFeliciano, S. G.; Kearsley, S. K. Designing targeted libraries with genetic algorithms. *J. Mol. Graph. Mod.* **2000**, *18*, 320-334.
- 159 Wright, T.; Gillet, V. J.; Green V. S. D.; Pickett, S. D. Optimizing the Size and Configuration of Combinatorial Libraries. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 381-390.
- 160 Zheng, W.; Cho, S. J.; Tropsha, A. Rational Combinatorial Library Design. 1. Focus-2D: A New Approach to the Design of Targeted Combinatorial Chemical Libraries. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 251-258.
- 161 Jamois, E. A.; Lin, C. T.; Waldman, M. Design of focused and restrained subsets from extremely large virtual libraries. *J. Mol. Graph. Mod.* **2003**, *22*, 141-149.
- 162 Agrafiotis, D. K.; Lobanov, V. S. Ultrafast algorithm for designing focused combinatorial arrays. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1030-1038.

- 163 Lobanov, V. S.; Agrafiotis, D. K. Stochastic Similarity Selections from Large Combinatorial Libraries. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 460-470.
- 164 Young, S. S.; Wang, M.; Gu, F. Design of Diverse and Focused Combinatorial Libraries Using an Alternating Algorithm. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1916-1921.
- 165 SYBYL, Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
- 166 Pascual, R.; Borrell, J. I.; Teixidó, J. Analysis of selection methodologies for combinatorial library design. *Mol. Divers.* **2003**, *6*(2), 121-133.
- 167 Gillet, V. J.; Willet, P.; Fleming, P.; Green, D. V. S. Designing focused libraries using MoSELECT. *J. Mol. Graph. Mod.* **2002**, *20*, 491-498.
- 168 Gutiérrez-de-Terán, H.; Lozano, J. J.; Segarra, V.; Sanz, F. Molecular Diversity Sample Generation on the Basis of Quantum-Mechanical Computations and Principal Component Analysis. *Combi. Chem. High Throughput Screen.* **2002**, *5*(1), 49-57.
- 169 Rusinko III, A.; Sheridan, R. P.; Nilakantan, R.; Haraki, K. S.; Bauman, N.; Venkataraghavan, R. Using CONCORD To Construct a Large Database of Three-Dimensional Coordinates from Connection Tables. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 251-255.
- 170 Sadowski, J.; Schwab, C. H.; Gasteiger, J. Automatic generation of 3D atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537-547.
- 171 Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity – Rapid access to atomic charges. *Tetrahedron.* **1980**, *36*, 3219-3288.
- 172 Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints For Determining Atom-Centered Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269-10280.
- 173 Bayly, C. I.; Jakalian, A. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623-1641.
- 174 Allinger, N. L. Conformational Analysis 130. MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms. *J. Am. Chem. Soc.* **1977**, *99*: 8127-8134.
- 175 Allinger, N. L.; Yuh, Y. H.; Lii, J-H. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551-8565.
- 176 Allinger, N. L.; Chen, K.; Lii, J. H. An Improved Force Field (MM4) for Saturated Hydrocarbons. *J. Comp. Chem.* **1996**, *17*, 642-668.
- 177 Clark, M.; Cramer, R. D.; Van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comp. Chem.* **1989**, *10*, 982-1012.
- 178 Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization and Performance of MMFF94. *J. Comp. Chem.* **1996**, *17*, 490-519.
- 179 Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. A.; Skiff; W. M. UFF, a Rule-Based Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024-10035.
- 180 Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.



- 181 Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, Jr.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
- 182 Brooks, B. R.; *et al.* CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comp. Chem.* **1983**, *4*, 187-217.
- 183 Hermans, J.; Berendsen, H. J.; Van Gunsteren, W. F.; Postma, J. P. A Consistent Empirical Potential for Water-Protein Interactions. *Biopolymers.* **1984**, *23(8)*, 1513-1518.
- 184 Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W.L. OPLS All-Atom Force Field for Carbohydrates. *J. Comp. Chem.* **1997**, *18*, 1955-1970.
- 185 Ponder, W.; Case, D. A. Force fields for protein simulations. *Adv. Prot. Chem.* **2003**, *66*, 27-85.
- 186 Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. K. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.
- 187 Case, D. A.; Pearlman D. A.; Caldwell J. W.; Cheatham III.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, L.; Seibel G. L., Singh, U. C.; Weiner, P. K.; Kollman, P. A. (2002). AMBER 7. University of California. San Francisco. (2004). AMBER 8. University of California. San Francisco.
- 188 Leach, A. R. Energy Minimisation and Related Methods for Exploring the Energy Surface. En *Molecular Modelling. Principles and Applications*. Prentice Hall: Great Britain, 2001; Chapter 5, pp 253-302.
- 189 Leach, A. R. Computer Simulation Methods. Molecular Dynamics Simulation Methods. En *Molecular Modelling. Principles and Applications*. Prentice Hall: Great Britain, 2001; Chapter 6, pp 303-409.
- 190 Rapaport, D. C. *The Art of Molecular Dynamics Simulation*. Cambridge University Press: Great Britain, 1995.
- 191 Verlet, L.; Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98-103.
- 192 Hockney, R. W. The potential calculation and some applications. En *Methods in Computational Physics*; Alder, B.; Fernbach, S.; Rotenberg, M, Ed.; Plasma Physics. Academic Press: New York/London, 1970; Vol 9, pp 136-211.
- 193 Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters. *J. Chem. Phys.* **1982**, *76*, 637-649.
- 194 Gear, C. W. *Numerical Initial Value Problems in Ordinary Differential Equations*. Prentice Hall: USA, 1971.
- 195 Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.
- 196 Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- 197 Andrea, T. A.; Swope, W. C.; Andersen, H. C. The role of long ranged forces in determining the structure and properties of liquid water. *J. Chem. Phys.* **1983**, *79*, 4576-4584.

- 198 Sagui, C.; Darden, T. A.; Molecular Dynamics Simulations of Biomolecules: Long-Range Electrostatic Effects. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155-179.
- 199 Darden, T. A.; York, D. M.; Pedersen, L. G. Particle mesh Ewald: an  $N \log .N/$  method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
- 200 Orozco, M.; Luque, F. J. Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems. *Chem. Rev.* **2000**, *100*, 4187-4225.
- 201 Luo, R.; Lu, Q. A Poisson-Boltzmann dynamics method with nonperiodic boundary condition. *J. Chem. Phys.* **2003**, *119(21)*, 11035-11047.
- 202 Leach, A. R. Four Challenges in Molecular Modelling. En *Molecular Modelling. Principles and Applications*. Prentice Hall: Great Britain, 2001; Chapter 11, pp 603-608.
- 203 Gilson, M. K.; Honig, B. Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies and Conformational Analysis. *Proteins.* **1998**, *4*, 7-18.
- 204 Bashford, D.; Gerwert, K. Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* **1992**, *224(2)*, 473-86.
- 205 Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. Electrostatics and Diffusion of Molecules in Solution: Simulations with the University of Houston Brownian Dynamics Program. *Comp. Phys. Comm.* **1991**, *62*, 187-197.
- 206 Delphi. Theory and Application. Accelrys website. [http://www.bio.unizh.ch/docu/acc\\_docs/doc/delphi/02\\_Theory.html](http://www.bio.unizh.ch/docu/acc_docs/doc/delphi/02_Theory.html). En línea (16/2/2006).
- 207 Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978-1988.
- 208 Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.
- 209 Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824-19839.
- 210 Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes With a Modified Generalized Born Model. *Proteins.* **2004**, *55*, 383-394.
- 211 Ajay; Murcko, M. A. Computational methods to predict binding free energy in ligand-receptor complexes. *J. Med. Chem.* **1995**, *38(26)*, 4953-4967.
- 212 Stouten, P. F. W.; Froemmel, C.; Nakamura, H.; Sander, C. An Effective Solvation Term Based on Atomic Occupancies for Use in Protein Simulations. *Mol. Simulat.* **1993**, *10*, 97-120.
- 213 Cole, J. C.; Nissink, W. M.; Taylor, R. Protein-Ligand Docking and Virtual Screening with GOLD. En *Virtual Screening in Drug Discovery*; Alvarez, J.; Shoichet, B., Ed.; Taylor & Francis Group. Florida, 2005; pp 379-415.
- 214 Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: 1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput Aided Mol. Des* **1997**, *11*, 425-445.
- 215 Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Virtual Screening Using Protein-Ligand Docking: Avoiding Artificial Enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793-806.

- 216 Pierce, A. C.; Sandretto, K. L.; Bermis, G. W. Kinase inhibitors and the case for C-H...O hydrogen bonds in protein-ligand binding. *Proteins*. **2002**, *49*: 567-576.
- 217 Connolly, M. Molecular Surfaces: A review. Accesible a través de <http://www.netsci.org/Science/Compchem/feature14.html>. En línea (1/3/06).
- 218 Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Cryst.* **1983**, *16*, 548-558.
- 219 Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217-230.
- 220 Wang, J.; Kang, K.; Kuntz, I. D.; Kollman, P. A. Hierarchical Database Screenings for HIV-1 Reverse Transcriptase Using a Pharmacophore Model, Rigid Docking, Solvation Docking, and MM-PB/SA. *J. Med. Chem.* **2005**, *48* (7), 2432-2444.
- 221 Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*(22), 5221-30.
- 222 Gohlke, H.; Case, D. A. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* **2004**, *25*(2), 238-250.
- 223 Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403-410. Accesible a través de <http://www.ncbi.nlm.nih.gov/BLAST/>. En línea (2/3/06).
- 224 Pearson, W. R.; Lipman, D. J. Improved Tools for Biological Sequence Comparison. *PNAS*. **1988**, *85*, 2444-2448. En línea (2/3/06). Accesible a través de <http://www.ebi.ac.uk/fasta33/#>.
- 225 Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M. C.; Estreicher, A.; Gasteiger, E.; Martín, M. J.; Michoud, K.; O'Donovan, C.; Phan, I.; Pilbout, S.; Schneider, M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic. Acids. Res.* **2003**, *31*, 365-370. Accesible a través de: <http://us.expasy.org/sprot/>. En línea (2/3/06).
- 226 Yongxing, C.; Zhang-Zhim H.; Robert, S.; Ledley, K.; Lewis, C.; Hans-Werner, M.; Barker. The Protein Information Resource: an integrated public resource of functional annotation of proteins. *Nucleic. Acids. Res.* **2002**, *30*, 35-37. Accesible a través de <http://pir.georgetown.edu/pirwww/>. En línea (2/3/06).
- 227 Bleasby, A. J.; Akrigg, D.; Attwood, T. K. OWL - A non-redundant, composite protein sequence database. *Nucleic Acids Res.* **1994**, *22*(17), 3574-3577. Accesible a través de <http://bioinf.man.ac.uk/dbbrowser/OWL/>. En línea (2/3/06).
- 228 Apweiler, R.; Bairoch, A.; Wu, C. H.; Barrer, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. L. UniProt: the Universal Protein knowledgebase. *Nucleic. Acids. Res.* **2004**, *32*: D115-D119.
- 229 Horn, F.; Weare, J.; Beukers, M. W.; Hörsch, S.; Bairoch, A.; Chen, W.; Edvardsen Ø.; Champagne, F.; Vriend, G. GPCRDB: an information system for G protein-coupled receptors. *Nucleic. Acids. Res.* **1998**, *26*(1), 277-281. Accesible a través de <http://www.gpcr.org/7tm/>. En línea (2/3/06).
- 230 Smith, C. M.; Shindyalov, I. N.; Veretnik, S.; Gribskov, M.; Taylor, S. S.; Ten Eyck, L. F.; Bourne, P. E. The protein kinase resource. *Trends. Biochem. Sci.* **1997**, *22*(11), 444-446. Accesible a través de <http://www.kinasenet.org/pkr/static.do?p=/content/references.html>. En línea (2/3/06).
- 231 Apweiler, R.; Bairoch, A.; Wu, C. H. Protein sequence databases. *Curr. Opin. Chem. Biol.* **2004**, *8*, 76-80.

- 232 Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*. **2000**, *289*(5480), 733-734.
- 233 Needleman, S.; Wunsch, C. A. General method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443-453.
- 234 Smith, T.F.; Waterman, M.S. Identification of Common Molecular Subsequences. *J. Mol. Biol.* **1981**, *147*, 195-197.
- 235 Altschul, S. F. Amino acid substitution matrices from an information theoretic perspective. *J. Mol. Biol.* **1991**, *219*, 555-565.
- 236 Jones, D. T.; Taylor, W. R.; Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *CABIOS*. **1992**, *8*, 275-282.
- 237 Gonnet, G. H.; Cohen, M. A.; Benner, S. A. Exhaustive matching of the entire protein sequence database. *Science*. **1992**, *256*, 1443-1445.
- 238 Henikoff, S.; Henikoff, J. G. Amino acid substitution matrices from protein blocks. *PNAS*. **1992**, *89*, 10915-10919.
- 239 Altschul, S. F. A protein alignment scoring system sensitive at all evolutionary distances. *J. Mol. Evol.* **1993**, *36*, 290-300.
- 240 Henikoff, S.; Henikoff, J.G. Performance Evaluation of amino acid substitution matrices. *Proteins*. **1993**, *17*(1), 49-61.
- 241 MacKerell, A.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*. **1998**, *102*, 3586-3616.
- 242 Sali, A.; Overington, J. Derivation of rules for comparative protein modeling from a database of protein structure alignments. *Protein Science*. **1994**, *3*, 1582-1596.
- 243 *Proteins Structure Prediction. Methods and Protocols*. Methods in Molecular Biology; Webster, D.M., Ed.; Humana Press: UK: 2000, pp 1-417.
- 244 Fidelis, K.; Stern, P.S.; Bacon, D.; Moulton, J. Comparison of systematic search and database methods for constructing segments of protein structure. *Protein. Eng.* **1994**, *7*, 953-960.
- 245 Sippl, M. J. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structure in globular proteins. *J. Mol. Biol.* **1990**, *213*, 859-883.
- 246 Crum-Rwon.; Frazer, A. Trans. R. Soc. Edinburgh, **1868-9**, *25*, 151-203.
- 247 Livingstone, D. J. The characterization of chemical structures using molecular properties. A survey. *J. Chem. Inf. Comput. Sci.* **2000**, *40*(2), 195-209.
- 248 Gutman, I.; Ruscic, B.; Trinajstić, N.; Wilcox, C. J. Graph Theory and Molecular Orbitals. XII. Acyclic Polyenes. *J. Chem. Phys.* **1975**, *62*, 3399-3405.
- 249 Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17-20.
- 250 Wiener, H. Correlation of Heat of Isomerization, and Differences in Heats of Vaporization of Isomers, among the Paraffin Hydrocarbons. *J. Am. Chem. Soc.* **1947**, *69*, 2636-2638.
- 251 Wiener, H. Influences of Interatomic Forces on Paraffin Properties. *J. Chem. Phys.* **1947**, *15*, 766-766.

- 252 Kier, L. B. A Shape Index from Molecular Graphs. *Quant. Struct.-Act. Relat.* **1985**, *4*, 109-116.
- 253 Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399-404.
- 254 Balaban, A. T. Topological Indices Based on Topological Distances in Molecular Graphs. *Pure & Appl.Chem.* **1983**, *55*, 199-206.
- 255 Petitjean, M. Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 331-337.
- 256 Randic, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609-6615.
- 257 Kier, L. B.; Hall, L. H. The Nature of Structure-Activity Relationships and their Relation to Molecular Connectivity. *Eur. J. Med. Chem.* **1977**, *12*, 307-312.
- 258 Kier, L. B.; Hall, L. H. Derivation of Significance of Valence Molecular Connectivity. *J.Pharm.Sci.* **1981**, *70*, 583-589.
- 259 Kier, L. B.; Hall, L. H. General Definition of Valence Delta-Values from Molecular Connectivity. *J. Pharm. Sci.* **1983**, *72*, 1170-1173.
- 260 Kier, L. B. Distinguishing Atom Differences in Molecular Graph Shape Index. *Quant. Struct.-Act. Relat.* **1986**, *5*, 7-12.
- 261 Kier, L. B. An Index of Molecular Flexibility from Kappa Shape Attributes. *Quant. Struct.-Act. Relat.* **1989**, *8*, 221-224.
- 262 Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39(5)*, 868-873.
- 263 Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three-dimensional Structure Directed Quantitative Structure-Activity Relationships. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565-577.
- 264 MOE LogP(Octanol/Water) Model. Labute, P. Source code in MOE/lib/svl/quasar.svl/q\_logp.svl, unpublished. 1998.
- 265 Labute, P. MOE Molar Refractivity Model. *unpublished*. 1998.
- 266 Stewart, J. J. P. MOSOL, MOPAC for Solid-State Physics. *Quant. Chem. Prog. Exch.* **1985**, *5*, 62-63.
- 267 Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714-3717.
- 268 Stanton, D. T.; Jurs, P. C. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationships studies. *Anal. Chem.* **1990**, *62*, 2323-2329.
- 269 Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure-Activity Studies: Definitions and Applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64-73.
- 270 Willett, P.; Winterman, V.; Bawden, D. Implementation of nearest neighbour searching in an online chemical structure search system. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 36-41.

- 271 Pepperrell, C. A.; Willett, P. Techniques for the Calculation of Three-Dimensional Structural Similarity Using Inter-Atomic Distances. *J. Comput Aided Mol. Des.* **1991**, *5*, 455-474.
- 272 Bemis, G. W.; Kuntz, I. D. A Fast and Efficient Method for 2D and 3D Molecular Shape Description. *J. Comput Aided Mol. Des.* **1992**, *6*, 607-628.
- 273 Bath, P. A.; Poirrette, A. R.; Willett, P.; Allen, F. H. Similarity Searching in Files of Three-Dimensional Chemical Structures: Comparison of Fragment-Based Measures of Shape Similarity. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 141-147.
- 274 Nilakantan, R.; Bauman, N.; Venkataraghavan, R. A New Method for Rapid Characterization of Molecular Shape: Applications in Drug Design. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 79-85.
- 275 Good, A. C.; Kuntz, I. D. Investigating the extension of pairwise distance pharmacophore measures to triplet-based descriptors. *J. Comput Aided Mol. Des.* **1995**, *9*(4), 373-379.
- 276 Sheridan, R. P.; Miller, M. D.; Underwood, D. J.; Kearsley, S. K. Chemical Similarity Using Geometric Pair Descriptors. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 128-136.
- 277 Pickett, S. D.; Mason, J. S.; McLay, I. M. Diversity Profiling and Design Using 3D Pharmacophores: Pharmacophore-Derived Queries (PDQ). *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1214-1223.
- 278 McGregor, M. J.; Muskal, S. M. Pharmacophore fingerprinting. 1. Application to QSAR and focused library design. *J. Chem. Inf. Comput. Sci.* **1999**, *39*(3), 569-74.
- 279 Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; Labaudiniere, R. F. New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* **1999**, *42*(17), 3251-3264.
- 280 Clark, R. D.; Fox, P. C.; Abrahamian, E. Using Pharmacophore Multiplet Fingerprints for Virtual High Throughput Screening. En *Virtual Screening in Drug Discovery*; Alvarez, J.; Shoichet, B., Ed.; Taylor & Francis Group. Florida, 2005; pp 207-225.
- 281 Makara, G. M. Measuring molecular similarity and diversity: total pharmacophore diversity. *J. Med. Chem.* **2001**, *44*(22), 3563-3571.
- 282 Horvath, D.; Jeandenans, C. Neighborhood behavior of in silico structural spaces with respect to in vitro activity spaces-a novel understanding of the molecular similarity principle in the context of multiple receptor binding profiles. *J. Chem. Inf. Comput. Sci.* **2003**, *43*(2), 680-690.
- 283 Fechner, U.; Franke, L.; Renner, S.; Schneider, P.; Schneider, G. Comparison of correlation vector methods for ligand-based similarity searching. *J. Comput Aided Mol. Des.* **2003**, *17*(10), 687-698.
- 284 Bush, B. L.; Sheridan, R. P. PATTY: A programmable atom type and language for automatic classification of atoms in molecular databases. *J. Chem. Inf. Comput. Sci.* **1993**, *33*(5), 756-762.
- 285 MOE. Función *ph4\_aType*.
- 286 Patel, Y.; Gillet, V. J.; Bravi, G.; Leach, A. R. A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP. *J. Comput Aided Mol. Des.* **2002**, *16*(8-9), 653-681.
- 287 Labute, P.; Williams, C.; Feher, M.; Sourial, E.; Schmidt, J. M. Flexible alignment of small molecules. *J. Med. Chem.* **2001**, *44*(10), 1483-1490.
- 288 Burkard, A. Methods for Data Analysis. En *Cheminformatics*; Gasteiger, J.; Engels, T., Ed.; Verlag GmbH & Co: Weinheim, 2003; pp 439-485.

- 289 Agrafiotis, D. K.; Lobanov, V. S. Nonlinear mapping networks. *J. Chem. Inf. Comput. Sci.* **2000**, *40*(6), 1356-1362.
- 290 Parrill, A. Introduction to Evolutionary Algorithms. En *Evolutionary Algorithms in Molecular Design*; Clark, D. E.; Mannhold, R.; Kubinyi, H.; Timmerman, H., Ed.; Wiley-VCH Verlag: Weinheim, 2000.
- 291 Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Optimization by Simulated Annealing. *Science*. **1983**, *220*, 671-680.
- 292 Goldberg, D. E. *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley.USA: 1989.
- 293 Pareto, V. Manual of Political Economy. 1906.
- 294 Handschuh, S.; Wagener, M.; Gasteiger, J. Superposition of Three-Dimensional Chemical Structures Allowing for Conformational Flexibility by a Hybrid Method. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 220-232.
- 295 Gillet, V. J.; Khatib, W.; Willett, P.; Fleming, P. J.; Green, D. V. Combinatorial library design using a multiobjective genetic algorithm. *J. Chem. Inf. Comput. Sci.* **2002**, *42*(2), 375-85.
- 296 Agrafiotis, D. K. Multiobjective optimization of combinatorial libraries. *J. Comput Aided Mol. Des.* **2002**, *16*, 335-356.
- 297 Nicolotti, O.; Gillet, V. J.; Fleming, P. J.; Green, D. V. S. Multiobjective Optimization in Quantitative Structure-Activity Relationships: Deriving Accurate and Interpretable QSARs. *J. Med. Chem.* **2002**, *45*(23), 5069-5080.
- 298 Brown, N.; McKay, B.; Gilardoni, F.; Gasteiger, J. A Graph-Based Genetic Algorithm and Its Application to the Multiobjective Evolution of Median Molecules. *J. Chem. Inf. Comput. Sci.* **2004**, *44*(3), 1079-1087.
- 299 Cottrell, S. J.; Gillet, V. J.; Taylor, R.; Wilton, D. J. Generation of multiple pharmacophore hypotheses using multiobjective optimisation techniques. *J. Comput Aided Mol. Des.* **2004**, *18*, 665-682.
- 300 Kochev, N.; Monev, V.; Bangov, I. Searching Chemical Structures. En *Chemoinformatics*; Gasteiger, J.; Engels, T., Ed.; Verlag GmbH & Co: Weinheim, 2003; pp 291-318.
- 301 Willet, P. Subset-Selection Methods For Chemical Databases. *Methods for Subset Selection*. En *Molecular Diversity in Drug Design*. Dean, P. M.; Lewis, R. A., Ed.; Kluwer Academic Publishers: The Netherlands, 1999; pp 115-140.
- 302 Pearlman, R. S.; Smith, K. M.; Deanda, F. Low-dimensional chemistry spaces: recent advances. Paper presented at the Cambridge Healthtech Institute conference "Chemoinformatics" held in Boston 15-16 June 1998.
- 303 Pascual, R. Tesis Doctoral. *Estudios sobre la metodología computacional para el diseño de bibliotecas combinatorias: Desarrollo del programa Pralins y aplicaciones en el campo del HIV*. IQS. Barcelona, 2003.
- 304 Holliday, J. D.; Ranade, S. S.; Willett, P. A Fast Algorithm For Selection Sets of Dissimilar Molecules From Large Chemical Databases. *Quant. Struct.-Act. Relat.* **1995**, *14*, 501-506.
- 305 Agrafiotis, D. K. Stochastic Algorithms for Maximizing Molecular Diversity. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 841-851.
- 306 Hassan, M.; Bielawski, J. P.; Hempel, J. C.; Waldman, M. Optimization and visualization of molecular diversity of combinatorial libraries. *Mol. Divers.* **1996**, *2*, 64-74.

- 307 Clark, R. D. OptiSim: An Extended Dissimilarity Selection Method for Finding Diverse Representative Subsets. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1181-1188.
- 308 Downs, G. M.; Barnard, J. M. Clustering Methods and Their Uses in Computational Chemistry. En *Reviews in Computational Chemistry*, Lipkowitz, K. B.; Boyd, D. B., Ed.; Wiley and Sons: New York, 2002; Vol 18, pp 1-40.
- 309 Cluster Algorithms. Methods and Data Quantlets. Accesible a través de <http://www.quantlet.com/mdstat/scripts/mva/htmlbook/mvahtmlnode80.html>. En línea (3/9/04).
- 310 Bravi, G.; Green, D. V.; Hann, M. M.; Leach, A. R. PLUMS: a program for the rapid optimization of focused libraries. *J. Chem. Inf. Comput. Sci.* **2000**, *40*(6), 1441-1448.
- 311 Young, S. S. Design of Diverse and Focused Combinatorial Libraries Using an Alternating Algorithm. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1916-1921.
- 312 Mason, J. S.; Pickett, S. D. Partition-based selection. *Perspect. Drug. Discovery. Des.* **1997**, *7*/8, 85-114.
- 313 Pearlman, R. S.; Wang, X. C.; Xu, Y.; Green, M. Novel methods for assessing and comparing the diversity of chemical libraries; 218th ACS meeting: New Orleans, August 22-26,1999.
- 314 Turner, D. B.; Tyrrell, S. M.; Willett, P. Rapid Quantification of Molecular Diversity for Selective Database Acquisition. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 18-22.
- 315 Rubin, G. M.; Yandell, M. D.; Wortman, J. R.; Gabor Miklos, G. L.; Nelson, C. R.; Hariharan, I. K.; Fortini, M. E.; Li, P.; W. Apweiler, R.; Fleischmann, W.; *et al.* Comparative genomics of the eukaryotes. *Science*. **2000**, *287*, 2204-2215.
- 316 Cohen, P. Protein kinases the major drug targets of the twenty first century? *Nat. Rev. Drug Discovery*. **2002**, *1*, 309-315.
- 317 Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science*. **2002**, *298*, 1912-1934.
- 318 Cell Signaling Technology. Accesible a través de <http://www.cellsignal.com/reference/kinase/index.asp>. En línea (1/4/06).
- 319 Evolutionary Bioinformatics and Sugen, Inc. Accesible a través de <http://kinase.com/human/kinome/>. En línea (8/4/06).
- 320 Hanks, S. K.; Quinn, A. M. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol.* **1991**, *200*, 38-62.
- 321 Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell*. **2000**, *103*, 211-225.
- 322 Pawson, T.; Gish, G. D. SH2 and SH3 domains: from structure to function. *Cell*. **1992**, *71*(3), 359-362.
- 323 Jones, S. M.; Kazlauskas, A. Growth Factor-Dependent Signaling and Cell Cycle Progression. *Chem. Rev.* **2001**, *101*, 2413-2423.
- 324 Voldborg, B. R.; Damstrup, L.; Spang-Thomsen, M.; Poulsen, H. S. Epidermal growth factor receptor (EGFR) and EGFR mutations, function and possible role in clinical trials. *Ann. Oncol.* **1997**, *12*, 1197-1206.
- 325 Platelet-derived growth factor receptor. *Cytokine Growth F. R.* **2004**, *15*(4), 195-289.
- 326 Fibroblast growth factor receptor. *Cytokine Growth F. R.* **2005**, *16*(2), 107-247.



- 327 Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K. Watson, J. D. *Biología Molecular de la célula*. Ediciones Omega: Barcelona, 2002; pp 966-968.
- 328 Porter, A. C.; Vaillancourt, R. R. Tyrosine kinase receptor-activated signal transduction pathways which lead to oncogenesis. *Oncogene*. **1998**, *17*, 1343-1352.
- 329 Kinase-Disease Associations. Accesible a través de [http://www.cellsignal.com/reference/kinase\\_disease.asp](http://www.cellsignal.com/reference/kinase_disease.asp). En línea (9/4/06).
- 330 Hubbard, S. R.; Till, J. H. Protein tyrosine kinase structure and function. *Annu. Rev. Biochem.* **2000**, *69*, 373-98.
- 331 Hanks, S. K.; Quinn, A. M.; Hunter, T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science*. **1988**, *241*(4861), 42-52.
- 332 Taylor, S. S.; Radzio-Andzelm, E. Three protein kinase structures define a common motif. *Structure*. **1994**, *2*(5), 345-55.
- 333 Knighton, D. R.; Zheng, J. H.; Ten Eyck, L. F.; Ashford, V. A.; Xuong, N. H.; Taylor, S. S.; Sowadski, J. M. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*. **1991**, *253*(5018), 407-414.
- 334 García-Echeverría, C. RPTKs: the new frontier of targeted anti-cancer agents. *DrugPlus International*. **2004**, *6*, 10-13.
- 335 García-Echeverria, C.; Traxler, P.; Evans, D. B. ATP site-directed competitive and irreversible inhibitors of protein kinases. *Med. Res. Rev.* **2000**, *20*(1), 28-57.
- 336 Bridges, A. Chemical Inhibitors of Protein Kinases. *Chem. Rev.* **2001**, *101*, 2541-2571.
- 337 Hennequin, L. F.; Thomas, A. P.; Johnstone, C.; Stokes, E. S. Pl inverted question marke PA, Lohmann, J. J.; Ogilvie, D. J.; Dukes, M.; Wedge, S. R.; Curwen, J. O.; Kendrew, J.; Lambert-van der Brempt, C. Design and structure-activity relationship of a new class of potent VEGF receptor tyrosine kinase inhibitors. *J. Med. Chem.* **1999**, *42*(26), 5369-89.
- 338 Smaill, J. B.; Palmer, B. D.; Rewcastle, G. F.; Denny, W. A.; McNamara, D. J.; Dobrusin, E. M.; Bridges, A. J.; Zhou, H.; Showalter, H. D.; Winters, R. T.; Leopold, W. R.; Fry, D. W.; Nelson, J. M.; Slintak, V.; Elliot, W. L.; Roberts, B. J.; Vincent, P. W.; Patmore, S. J. Tyrosine kinase inhibitors. 15. 4-(Phenylamino)quinazoline and 4-(phenylamino)pyrido[d]pyrimidine acrylamides as irreversible inhibitors of the ATP binding site of the epidermal growth factor receptor. *J Med Chem.* **1999**, *42*, 1803-1815.
- 339 Zimmermann, J.; Buchdunger, E.; Mett, H.; Meyer, T.; Lydon, N. B. Potent and selective inhibitors of the Abl-kinase phenylamino-pyrimidine (PAP) derivatives. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 187-192.
- 340 Rewcastle, G. F.; Palmer, B. D.; Thompson, A. M.; Bridges, A. J.; Cody, D. R.; Zhou, H. M.; Fry, D. W.; McMichael, A.; Denny, W. A. Tyrosine kinase inhibitors. 10. Isomeric 4-[(3-bromophenyl)amino]pyrido[d]-pyrimidines are potent ATP binding site inhibitors of the tyrosine kinase function of the epidermal growth factor receptor. *J. Med. Chem.* **1996**, *39*, 1823-1835.
- 341 Dahring, T. K.; Lu, G. H.; Hamby, J. M.; Batley, B. L.; Kraker, A. J.; Panek, R. L. Inhibition of growth factor-mediated tyrosine phosphorylation in vascular smooth muscle by PD 089828, a new synthetic protein tyrosine kinase inhibitor. *J. Pharmacol. Exp. Ther.* **1997**, *281*(3), 1446-1456.

- 342 Klutchko, S. R.; Hamby, J. M.; Boschelli, D. H.; Wu, Z.; Kraker, A. J.; Amar, A.; Hartl, B. G.; *et al.* 2-Substituted aminopyrido[2,3-*d*]pyrimidin-7(8*H*)-ones. structure-activity relationships against selected tyrosine kinases and in vitro and in vivo anticancer activity. *J Med Chem.* **1998**, *41*, 3276-3292.
- 343 Connolly, C. J. C.; Hamby, J. M.; Schroeder, M. C.; Barvian, M.; Lu, G. H.; Panek, R. L.; Amar, A.; Shen, C.; Kraker, A. J.; Fry, D. W.; Klohs, W. D.; Doherty, A. M. Discovery and structure-activity studies of a novel series of pyrido[2,3-*d*]pyrimidine tyrosine kinase inhibitors. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2415-2420.
- 344 Hamby, J. M.; Connolly, C. J.; Schroeder, M. C.; Winters, R. T.; Showalter, H. D.; Panek, R. L.; Major, T. C.; Olsewski, B.; Ryan, M. J.; Dahring, T.; Lu, G. H.; Keiser, J.; Amar, A.; Shen, C.; Kraker, A. J.; Slintak, V.; Nelson, J. M.; Fry, D. W.; Bradford, L.; Hallak, H.; Doherty, A. M. Structure-activity relationships for a novel series of pyrido[2,3-*d*]pyrimidine tyrosine kinase inhibitors. *J. Med. Chem.* **1997**, *40*, 2296-2303.
- 345 Schroeder, M. C.; Hamby, J. M.; Connolly, C. J.; Grohar, P. J.; Winters, R. T.; Barvian, M. R.; Moore, C. W.; Boushelle, S. L.; Crean, S. M.; Kraker, A. J.; Driscoll, D. L.; Vincent, P. W.; Elliott, W. L.; Lu, G. H.; Batley, B. L.; Dahring, T. K.; Major, T. C.; Panek, R. L.; Doherty, A. M.; Showalter, H. D. Soluble 2-substituted aminopyrido[2,3-*d*]pyrimidin-7-yl ureas. Structure-activity relationships against selected tyrosine kinases and exploration of in vitro and in vivo anticancer activity. *J. Med. Chem.* **2001**, *44*(12), 1915-26.
- 346 Trumpp-Kallmeyer, S.; Rubin, J. R.; Humblet, C.; Hamby, J. M.; Showalter, H. D. H. Development of a binding model of protein tyrosine kinase for substituted pyrido[2,3-*d*]pyrimidine inhibitors. *J. Med. Chem.* **1998**, *41*, 1752-1763.
- 347 Boschelli, D. H.; Wu, Z.; Klutchko, S. R.; Showalter, H. D.; Hamby, J. M.; Lu, G. H.; Major, T. C.; Dahring, T.; Batley, B.; Panek, R. L.; Keiser, J.; Hartl, B. G.; Kraker, A. J.; Klohs, W. D.; Roberts, B. J.; Patmore, S.; Elliott, W. L.; Steinkampf, R.; Bradford, L.; Hallak, H.; Doherty, A. M. Synthesis and tyrosine kinase inhibitory activity of a series of 2-amino-8*H*-pyrido[2,3-*d*]pyrimidines: identification of potent, selective platelet-derived growth factor receptor tyrosine kinase inhibitors. *J. Med. Chem.* **1998**, *41*, 4365-4377.
- 348 Furet, P.; Caravatti, G.; Lydon, N.; Priestle, J. P.; Sowadski, J. M.; Trinks, U.; Traxler, P. Modelling study of protein kinase inhibitors: binding mode of staurosporine and origin of the selectivity of CGP 52411. *J. Comput Aided Mol. Des.* **1995**, *9*, 465-472.
- 349 Traxler, P.; Furet, P.; Mett, H.; Buchdunger, E.; Meyer, T.; Lydon, N. 4-(Phenylamino)pyrrolopyrimidines: potent and selective, ATP site directed inhibitors of the EGF-receptor protein tyrosine kinase. *J. Med. Chem.* **1996**, *39*, 2285-2292.
- 350 Henry, J. R.; Rupert, K. C.; Dodd, J. H.; Turchi, I. J.; Wadsworth, S. A.; Cavender, D. E.; Fahmy, B.; Olini, G. C.; Davis, J. E.; Pellegrino-Gensey, J. L.; Schafer, P. H.; Siekierka, J. J. 6-Amino-2-(4-fluorophenyl)-4-methoxy-3-(4-pyridyl)-1*H*-pyrrolo[2,3-*b*]pyridine (RWJ 68354): a potent and selective p38 kinase inhibitor. *J. Med. Chem.* **1998**, *41*(22), 4196-4198.
- 351 Hanke, J. H.; Gardner, J. P.; Dow, R. L.; Changelian, P. S.; Brissette, W. H.; Weringer, E. J.; Pollock, B. A.; Connelly, P. A. Discovery of a Novel, Potent, and Src Family-selective Tyrosine Kinase Inhibitor. *J. Biol. Chem.* **1996**, *271*, 695-701.
- 352 Traxler, P.; Bold, G.; Frei, J.; Lang, M.; Lydon, N.; Mett, H.; Buchdunger, E.; Meyer, T.; Mueller, M.; Furet, P. Use of a pharmacophore model for the design of EGF-R tyrosine kinase inhibitors: 4-(phenylamino)pyrazolo[3,4-*d*]pyrimidines. *J. Med. Chem.* **1997**, *40*(22), 3601-3616.
- 353 Sun, L.; Tran, F.; App, H.; Hirth, P.; McMahon, G.; Tang, C. Synthesis and biological evaluations of 3-substituted indolin-2-ones: a novel class of tyrosine kinase inhibitors that exhibit selectivity toward particular receptor tyrosine kinases. *J. Med. Chem.* **1998**, *41*, 2588-2603.

- 354 Sun, L.; Tran, N.; Liang, C.; Tang, F.; Rice, A.; Schreck, R.; Waltz, K.; Shawver, L. K.; McMahon, G.; Tang, C. Design, synthesis, and evaluations of substituted 3-[(3- or 4-carboxyethylpyrrol-2-yl)methylidene]indolin-2-ones as inhibitors of VEGF, FGF, and PDGF receptor tyrosine kinases. *J. Med. Chem.* **1999**, *42*(25), 5120-5130.
- 355 Sun, L.; Tran, N.; Liang, C.; Hubbard, S.; Tang, F.; Lipson, K.; Schreck, R.; Zhou, Y.; McMahon, G.; Tang, C. Identification of substituted 3-[(4,5,6, 7-tetrahydro-1*H*-indol-2-yl)methylene]-1,3-dihydroindol-2-ones as growth factor receptor inhibitors for VEGF-R2 (Flk-1/KDR), FGF-R1, and PDGF-Rbeta tyrosine kinases. *J. Med. Chem.* **2000**, *43*(14), 2655-2663.
- 356 Liverton, N. J.; Butcher, J. W.; Claiborne, C. F.; Claremon, D. A.; Libby, B. E.; Nguyen, K. T.; Pitzenberger, S. M.; Selnick, H. G.; Smith, G. R.; Tebben, A.; Vacca, J. P.; Varga, S. L.; Agarwal, L.; Dancheck, K.; Forsyth, A. J.; Fletcher, D. S.; Frantz, B.; Hanlon, W. A.; Harper, C. F.; Hofsess, S. J.; Kostura, M.; Lin, J.; Luell, S.; O'Neill, E. A.; O'Keefe, S. J.; *et al.* Design and synthesis of potent, selective, and orally bioavailable tetrasubstituted imidazole inhibitors of p38 mitogen-activated protein kinase. *J. Med. Chem.* **1999**, *42*, 2180-2190.
- 357 Palmer, B. D.; Smaill, J. B.; Boyd, M.; Boschelli, D. H.; Doherty, A. M.; Hamby, J. M.; Khatana, S. S.; Kramer, J. B.; Kraker, A.J.; Panek, R. L.; Lu, G. H.; Dahring, T. K.; Winters, R. T.; Showalter, H. D. H.; Denny, W. A. Structure-activity relationships for 1-phenylbenzimidazoles as selective ATP site inhibitors of the platelet-derived growth factor receptor. *J. Med. Chem.* **1998**, *41*, 5457-5465.
- 358 Palmer, B. D.; Kraker, A. J.; Hartl, B. G.; Panopoulos, A. D.; Panek, R. L.; Batley, B. L.; Lu, G. H.; Trumpp-Kallmeyer, S.; Hollis Showalter, H. D.; Denny, W. A. Structure-activity relationships for 5-substituted 1-phenylbenzimidazoles as selective inhibitors of the platelet-derived growth factor receptor. *J. Med. Chem.* **1999**, *42*(13), 2373-82.
- 359 Thompson, A. M.; Rewcastle, G. W.; Boushelle, S. L.; Hartl, B. G.; Kraker, A. J.; Lu, G. H.; Batley, B. L.; Panek, R. L.; Showalter, H. D.; Denny, W. A. Synthesis and structure-activity relationships of 7-substituted 3-(2, 6-dichlorophenyl)-1,6-naphthyridin-2(1*H*)-ones as selective inhibitors of pp60(c-src). *J. Med. Chem.* **2000**, *43*(16), 3134-3147.
- 360 McGovern, S. L.; Shoichet, B. K. Kinase inhibitors: not just for kinases anymore. *J. Med. Chem.* **2003**, *46*(8), 1478-1483.
- 361 Mont, N. Tesis Doctoral. *Diseño y síntesis combinatoria de sistemas heterocíclicos como inhibidores potenciales de tirosina quinasas*. IQS. Barcelona, 2005.
- 362 Mont, N.; Teixidó, J.; Borrell, J. I.; Kappe, O. A Three-Component Synthesis of Pyrido[2,3-*d*]pyrimidines. *Tetrahedron Lett.* **2003**, *44*, 5385-5387.
- 363 Mont, N.; Teixidó, J.; Kappe, C. O.; Borrell, J. I. A One-Pot Microwave-Assisted Synthesis of Pyrido[2,3-*d*]pyrimidines. *Mol. Divers.* **2003**, *7*, 153-159.
- 364 Scifinder@Scholar 2003. CAS.
- 365 ViewerPro. Accelrys, Inc. 9685 Scranton Road Sand Diego, CA 92121-3752. Accesible a través de la dirección <http://www.accelrys.com/products/dstudio/viewerpro/>. En línea (13/05/06).
- 366 ChemFinder 6.0. CambridgeSoft Corporation, 100 CambridgePark Drive, Cambridge, MA 02140 USA.
- 367 Rishton, G. M. Reactive compounds an *in vitro* false positives in HTS. *DDT.* **1997**, *2*, 382-384.
- 368 Waszkowycz, B.; Perkins, T. D. J.; Sykes, R. A.; Li, J. Large-scale Virtual Screening for Discovering Leads in the Postgenomic Era. *IBM Systems Journal* **2001**, *40*, 360-376.

- 369 Lyne, P. D.; Kenny, P. W.; Cosgrove, D. A.; Deng, C.; Zabludoff, S.; Wendoloski, J. J.; Ashwell, S. Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge-based virtual screening. *J. Med. Chem.* **2004**, *47*(8), 1962-1968.
- 370 Aronov, A. M.; Murcko, M. A. Toward a pharmacophore for kinase frequent hitters. *J. Med. Chem.* **2004**, *47*(23), 5616-5619.
- 371 Mohammadi, M.; Froum, S.; Hamby, J. M.; Schroeder, M. C.; Panek, R. L.; Lu, G. H.; Eliseenkova, A. V.; Green D.; Schlessinger, J.; Hubbard, S. R. Crystal Structure of an Angiogenesis Inhibitor Bound to the Fgf Receptor Tyrosine Kinase Domain. *Embo J.* **1998**, *17*, 5896-5904.
- 372 Hennequin, L. F.; Stokes, E. S. E.; Thomas, A. P.; Craig, J.; Plé, P. A.; Ogilvie, D. J. Dukes, M.; Wedge, S. R.; Kendrew, J. Curwen, J. O. Novel 4-Anilinoquinazolines with C-7 Basic Side Chains: Design and Structure Activity Relationship of a Series of Potent, Orally Active, VEGF Receptor Tyrosine Kinase Inhibitors. *J. Med. Chem* **2002**, *45*, 1300-1312.
- 373 Manetti, F. Small-molecule Inhibitors of Fibroblast Growth Factor Receptor (FGFR) Tyrosine Kinases (TK). *Curr. Pharm. Design.* **2003**, *9*, 567-581.
- 374 Schneider, P.; Schneider, G. Collection of Bioactive Reference Compounds for Focused Library Design. *QSAR & Comb. Sci.* **2003**, *22* (7), 713-718.
- 375 Mohammadi, M.; McMahon, G.; Sun, L.; Tang, C.; Hirth, P.; Yeh, B. K.; Hubbard, S. R.; Schlessinger, J. Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. *Science.* **1997**, *276*, 995-960.
- 376 Güner, O. F.; Henry, D. R. Metric for Analyzing Hit Lists and Pharmacophores. En *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: La Jolla, 2000, Chapter 11, pp 195-212.
- 377 Schneider, G.; Grabowski, T. Comunicación personal. 2005.
- 378 Daveu, C.; Bureau, R.; Baglin, I.; Prunier, H.; Lancelot, J. C.; Rault, S. Definition of a Pharmacophore for Partial Agonists of Serotonin 5-HT<sub>3</sub> Receptors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 362-369.
- 379 Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209-225.
- 380 Rockey, W. M.; Elcock, A. H. Rapid Computational Identification of the Targets of Protein Kinase Inhibitors. *J. Med. Chem.* **2005**, *48*, 4138-4152.
- 381 Diller, D. J.; Li, R. X. Kinases, homology models, and high throughput docking. *J. Med. Chem.* **2003**, *46*, 4638-4647.
- 382 Honma, T.; Hayashi, K.; Aoyama, T.; Hashimoto, N.; Machida, T.; Fukasawa, K.; Iwama, T.; Ikeura, C.; Ikuta, M.; Suzuki-Takahashi, I.; Iwasawa, Y.; Hayama, T.; Nishimura, S.; Morishima, H. Structure-based generation of a new class of potent CDK4 inhibitors: New *de novo* design strategy and library design. *J. Med. Chem.* **2001**, *44*, 4615-4627.
- 383 Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759-4767.
- 384 Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287-2303.
- 385 Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032-3047.

- 386 Kellenberger, E.; Rodrigo, J.; Muller, P.; *et al.* Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins*. **2004**, *57*(2), 225-242.
- 387 Perez, C.; Ortiz, A. R. Evaluation of docking functions for protein-ligand docking. *J. Med. Chem.* **2001**, *44*, 3768-3785.
- 388 Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of Docking Performance: Comparative Data on Docking Algorithms. *J. Med. Chem.* **2004**, *47*, 558-565.
- 389 Warren, G. L.; Andrews, C. V.; Capelli, A.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, F.; Senger, S.; Tedesco, G.; Wall, I. A.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.*, in press. (early view article).
- 390 Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, *42*, 5100-5109.
- 391 Feher, M. Consensus scoring for protein-ligand interactions. *DDT*, **2006**, *11*(9-10), 421-428.
- 392 Wang, R.; Wang, S. How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inf. Comput. Sci.* **2001**, *41*(5), 1422-1426.
- 393 Murray, C. W.; Baxter, C. A.; Frenkel, A. D. The sensitivity of the results of molecular docking to induced-fit effects: application to thrombin, thermolysin and neuraminidase. *J. Comput Aided Mol. Des.* **1999**, *13*(6), 547-562.
394. Birch, L.; Murray, C. W.; Hartshorn, M. J.; *et al.* Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. *J. Comput Aided Mol. Des.* **2002**, *16*(12), 855-869.
- 395 Erickson, J. A. ; Jalaie, M. ; Robertson, D. H. ; *et al.* Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47*(1), 45-55.
- 396 Thomas, M. P.; McInnes, C.; Fischerr, P.; Protein Structures in Virtual Screening: A Case Study with CDK2. *J. Med. Chem.* **2006**, *49*(1), 92-104.
- 397 McGovern, S. L.; Shoichet, B. K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*(14), 2895-2907.
- 398 Cavasotto, C. N.; Orry, A. J.; Abagyan, R. A. The Challenge of Considering Receptor Flexibility in Ligand Docking and Virtual Screening. *Curr. Comput. Aided Drug Des.* **2005**, *1*(4), 423-440.
- 399 Buzko, O. V.; Bishop, A. C.; Shokat, K. M. Modified AutoDock for accurate docking of protein kinase inhibitors. *J. Comput Aided Mol. Des.* **2002**, *16*, 113-127.
- 400 The Scripps Research Institute, La Jolla. Autodock: Parameters. Accesible a través de <http://www.scripps.edu/mb/olson/doc/autodock/parameters.html>. En línea (16/04/2006).
- 401 Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* **2002**, *277*(48), 46265-46272.
- 402 Wood, E. R.; Truesdale, A. T.; McDonald, O. B.; Yuan, D.; Hassell, A.; Dickerson, S. H.; Ellis, B.; Pennisi, C.; Horne, E.; Lackey, K.; Alligood, K. J.; Rusnak, D. W.; Gilmer, T. M.; Shewchuk, L. A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* **2004**, *64*(18), 6652-6659.

- 403 Zhang, X.; Gureasko, J.; Shen, K.; Cole, P. A.; Kuriyan, J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell*. **2006**, *125*(6), 1137-1149.
- 404 Wissner, A.; Berger, D. M.; Boschelli, D. H.; Floyd, M. B. Jr.; Greenberger, L. M.; Gruber, B. C.; Johnson, B. D.; Mamuya, N.; Nilakantan, R.; Reich, M. F.; Shen, R.; Tsou, H. R.; Upeslaciis, E.; Wang, Y. F.; Wu, B.; Ye, F.; Zhang, N. 4-Anilino-6,7-dialkoxyquinoline-3-carbonitrile inhibitors of epidermal growth factor receptor kinase and their bioisosteric relationship to the 4-anilino-6,7-dialkoxyquinazoline inhibitors. *J. Med. Chem.* **2000**, *43*(17), 3244-3256.
- 405 Hou, T.; Zhu, L.; Chen, L.; Xu, X. Mapping the binding site of a large set of quinazoline type EGF-R inhibitors using molecular field analyses and molecular docking studies. *J. Chem. Inf. Comput. Sci.* **2003**, *43*(1), 273-287.
- 406 Aparna, V.; Rambabu, G.; Panigrahi, S. K.; Sarma, J. A.; Desiraju, G. R. Virtual screening of 4-anilinoquinazoline analogues as EGFR kinase inhibitors: importance of hydrogen bonds in the evaluation of poses and scoring functions. *J. Chem. Inf. Model.* **2005**, *45*(3), 725-738.
- 407 AUTODOCK mailing list. ADL: Modelling Water. Morris, G. Accesible a través de <http://mgldev.scripps.edu/pipermail/autodock/2004-September/000367.html>. En línea (13/7/06).
- 408 Claesson-Welsh, L.; Eriksson, A.; Moren, A.; Severinsson, L.; Ek, B.; Ostman, A.; Betsholtz, C.; Heldin, C. H. cDNA cloning and expression of a human platelet-derived growth factor (PDGF) receptor specific for B-chain-containing PDGF molecules. *Mol. Cell. Biol.* **1988**, *8*(8), 3476-3486.
- 409 Yarden, Y.; Escobedo, J. A.; Kuang, W. J.; Yang-Feng, T. L.; Daniel, T. O.; Tremble, P. M.; Chen, E. Y.; Ando, M. E.; Harkins, R. N.; Francke, U.; *et al.* Structure of the receptor for platelet-derived growth factor helps define a family of closely related growth factor receptors. *Nature*. **1986**, *323*(6085), 226-232.
- 410 Bohmer, F. D.; Karagyozov, L.; Uecker, A.; Serve, H.; Botzki, A.; Mahboobi, S.; Dove, S. A single amino acid exchange inverts susceptibility of related receptor tyrosine kinases for the ATP site inhibitor STI-571. *J. Biol. Chem.* **2003**, *278*(7), 5148-5155.
- 411 Mol, C. D.; Dougan, D. R.; Schneider, T. R.; Skene, R. J.; Kraus, M. L.; Scheibe, D. N.; Snell, G. P.; Zou, H.; Sang, B. C.; Wilson, K. P. Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J. Biol. Chem.* **2004**, *279*(30), 31655-31663.
- 412 Mol, C. D.; Lim, K. B.; Sridhar, V.; Zou, H.; Chien, E. Y.; Sang, B. C.; Nowakowski, J.; Kassel, D. B.; Cronin, C. N.; McRee, D. E. Structure of a c-kit product complex reveals the basis for kinase transactivation. *J. Biol. Chem.* **2003**, *278*(34), 31461-31464.
- 413 Griffith, J.; Black, J.; Faerman, C.; Swenson, L.; Wynn, M.; Lu, F.; Lippke, J.; Saxena, K. The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. *Mol. Cell.* **2004**, *13*(2), 169-178.
- 414 McTigue, M. A.; Wickersham, J. A.; Pinko, C.; Showalter, R. E.; Parast, C. V.; Tempczyk-Russell, A.; Gehring, M. R.; Mroczkowski, B.; Kan, C. C.; Villafranca, J. E.; Appelt, K. Crystal structure of the kinase domain of human vascular endothelial growth factor receptor 2: a key enzyme in angiogenesis. *Structure*. **1999**, *7*(3), 319-30.
- 415 Mohammadi, M.; Schlessinger, J.; Hubbard, S. R. Structure of the FGF receptor tyrosine kinase domain reveals a novel autoinhibitory mechanism. *Cell*. **1996**, *86*(4), 577-87.
- 416 Schiering, N.; Knapp, S.; Marconi, M.; Flocco, M. M.; Cui, J.; Perego, R.; Rusconi, L.; Cristiani, C. Crystal structure of the tyrosine kinase domain of the hepatocyte growth factor receptor c-Met and its complex with the microbial alkaloid K-252a. *PNAS*. **2003**, *100*(22), 12654-12659.

- 417 Shewchuk, L. M.; Hassell, A. M.; Ellis, B.; Holmes, W. D.; Davis, R.; Horne, E. L.; Kadwell, S. H.; McKee, D. D.; Moore, J. T. Structure of the Tie2 RTK domain: self-inhibition by the nucleotide binding loop, activation loop, and C-terminal tail. *Structure*. **2000**, *8(11)*, 1105-1113.
- 418 Munshi, S.; Hall, D. L.; Kornienko, M.; Darke, P. L.; Kuo, L. C.; Structure of apo, unactivated insulin-like growth factor-1 receptor kinase at 1.5 Å resolution. *Acta Crystallogr D*. **2003**, *59*, 1725-1730.
- 419 Pautsch, A.; Zoepfel, A.; Ahorn, H.; Spevak, W.; Hauptmann, R.; Nar, H. Crystal structure of bisphosphorylated IGF-1 receptor kinase: insight into domain movements upon kinase activation. *Structure*. **2001**, *9(10)*, 955-965.
- 420 Munshi, S.; Kornienko, M.; Hall, D. L.; Reid, J. C.; Waxman, L.; Stirdivant, S. M.; Darke, P. L.; Kuo, L. C. Crystal structure of the Apo, unactivated insulin-like growth factor-1 receptor kinase. Implication for inhibitor specificity. *J. Biol. Chem.* **2002**, *277(41)*, 38797-38802.
- 421 Schindler, T.; Sicheri, F.; Pico, A.; Gazit, A.; Levitzki, A.; Kuriyan, J. Crystal structure of Hck in complex with a Src family-selective tyrosine kinase inhibitor. *Mol. Cell*. **1999**, *3*, 639-664.
- 422 Nagar, B.; Hantschel, O.; Young, M. A.; Scheffzek, K.; Veach, D.; Bornmann, W.; Clarkson, B.; Superti-Furga, G.; Kuriyan, J. Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell*. **2003**, *112(6)*, 859-71.
- 423 Nagar, B.; Bornmann, W. G.; Pellicena, P.; Schindler, T.; Veach, D. R.; Miller, W. T.; Clarkson, B.; Kuriyan, J. Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res*. **2002**, *62(15)*, 4236-4243.
- 424 Schindler, T.; Bornmann, W.; Pellicena, P.; Miller, W. T.; Clarkson, B.; Kuriyan, J. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science*. **2000**, *289(5486)*, 1938-1942.
- 425 Nowakowski, J.; Cronin, C. N.; McRee, D. E.; Knuth, M. W.; Nelson, C. G.; Pavletich, N. P.; Rogers, J.; Sang, B. C.; Scheibe, D. N.; Swanson, R. V.; Thompson, D. A. Structures of the cancer-related Aurora-A, FAK, and EphA2 protein kinases from nanovolume crystallography. *Structure*. **2002**, *10(12)*, 1659-1667.
- 426 Till, J. H.; Becerra, M.; Watty, A.; Lu, Y.; Ma, Y.; Neubert, T. A.; Burden, S. J.; Hubbard, S. R. Crystal structure of the MuSK tyrosine kinase: insights into receptor autoregulation. *Structure*. **2002**, *10(9)*, 1187-1196.
- 427 Lamers, M. B.; Antson, A. A.; Hubbard, R. E.; Scott, R. K.; Williams, D. H. Structure of the protein tyrosine kinase domain of C-terminal Src kinase (CSK) in complex with staurosporine. *J. Mol. Biol.* **1999**, *285(2)*, 713-25.
- 428 Favellyukis, S.; Till, J. H.; Hubbard, S. R.; Miller, W. T. Structure and autoregulation of the insulin-like growth factor 1 receptor kinase. *Nat. Struct. Biol.* **2001**, *8(12)*, 1058-1063.
- 429 Till, J. H.; Ablooglu, A. J.; Frankel, M.; Bishop, S. M.; Kohanski, R. A.; Hubbard, S. R. Crystallographic and solution studies of an activation loop mutant of the insulin receptor tyrosine kinase: insights into kinase mechanism. *J. Biol. Chem.* **2001**, *276(13)*, 10049-10055.
- 430 Hubbard, S. R.; Wei, L.; Ellis, L.; Hendrickson, W. A. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature*. **1994**, *372(6508)*, 746-754.
- 431 Ogawa, A.; Takayama, Y.; Sakai, H.; Chong, K. T.; Takeuchi, S.; Nakagawa, A.; Nada, S.; Okada, M.; Tsukihara, T. Structure of the carboxyl-terminal Src kinase, Csk. Structure of the carboxyl-terminal Src kinase, Csk. *J. Biol. Chem.* **2002**, *277(17)*, 14351-14354.

- 432 Zhu, X.; Kim, J. L.; Newcomb, J. R.; Rose, P. E.; Stover, D. R.; Toledo, L. M.; Zhao, H.; Morgenstern, K. A. Structural analysis of the lymphocyte-specific kinase Lck in complex with non-selective and Src family selective kinase inhibitors. *Structure*. **1999**, *7*(6), 651-661.
- 433 Yamaguchi, H.; Hendrickson, W. A. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature*. **1996**, *384*(6608), 484-489.
- 434 Li, S.; Covino, N. D.; Stein, E. G.; Till, J. H.; Hubbard, S. R. Structural and biochemical evidence for an autoinhibitory role for tyrosine 984 in the juxtamembrane region of the insulin receptor. *J. Biol. Chem.* **2003**, *278*(28), 26007-26014.
- 435 Hubbard, S. R. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *EMBO J.* **1997**, *16*(18), 5572-5581.
- 436 Hu, J.; Liu, J.; Ghirlando, R.; Saltiel, A. R.; Hubbard, S. R. Structural Basis for Recruitment of the Adapter Protein APS to the Activated Insulin Receptor *Mol. Cell.* **2003**, *12*, 1379-1389.
- 437 Wybenga-Groot, L. E.; Baskin, B.; Ong, S. H.; Tong, J.; Pawson, T.; Sicheri, F. Structural basis for autoinhibition of the Ephb2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region. *Cell*. **2001**, *106*(6), 745-757.
- 438 Mao, C.; Zhou, M.; Uckun, F. M. Crystal structure of Bruton's tyrosine kinase domain suggests a novel pathway for activation and provides insights into the molecular basis of X-linked agammaglobulinemia. *J. Biol. Chem.* **2001**, *276*(44), 41435-41443.
- 439 Witucki, L. A.; Huang, X.; Shah, K.; Liu, Y.; Kyin, S.; Eck, M. J.; Shokat, K. M. Mutant tyrosine kinases with unnatural nucleotide specificity retain the structure and phospho-acceptor specificity of the wild-type enzyme. *Chem. Biol.* **2002**, *9*(1), 25-33.
- 440 Xu, W.; Harrison, S. C.; Eck, M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature*. **1997**, *385*(6617), 595-602.
- 441 Williams, J. C.; Weijland, A.; Gonfloni, S.; Thompson, A.; Courtneidge, S. A.; Superti-Furga, G.; Wierenga, R. K. The 2.35 Å crystal structure of the inactivated form of chicken Src: a dynamic molecule with multiple regulatory interactions. *J. Mol. Biol.* **1997**, *274*(5), 757-775.
- 442 Foster, R.; Griffith, R.; Ferrao, P.; Ashman, L. Molecular basis of the constitutive activity and STI571 resistance of Asp816Val mutant KIT receptor tyrosine kinase. *J. Mol. Graph. Model.* **2004**, *23*(2), 139-152.
- 443 Chuaqui, C.; Deng, Z.; Singh, J. Interaction Profiles of Protein Kinase-Inhibitor Complexes and Their Application to Virtual Screening. *J. Med. Chem.* **2005**, *48*, 121-133.
- 444 Kelly, M. D.; Mancera, R. L. Expanded interaction fingerprint method for analyzing ligand binding modes in docking and structure-based drug design. *J. Chem. Inf. Comput. Sci.* **2004**, *44*(6), 1942-1951.
- 445 Mpamhanga, C. P.; Chen, B.; McLay, I. M.; Willett, P. Knowledge-based interaction fingerprint scoring: a simple method for improving the effectiveness of fast scoring functions. *J. Chem. Inf. Model.* **2006**, *46*(2), 686-698.
- 446 Stickle, D.F.; Rose, G. Hydrogen Bonding and Globular Proteins. *J. Mol. Biol.* **1992**, *226*, 1143-1159.
- 447 Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell*. **1998**, *95*(7), 927-937.



- 448 Das, K.; Ding, J.; Hsiou, Y.; Clark, A. D. Jr.; Moereels, H.; Koymans, L.; Andries, K.; Pauwels, R.; Janssen, P. A.; Boyer, P. L.; Clark, P.; Smith, R. H. Jr.; Kroeger Smith, M. B.; Michejda, C. J.; Hughes, S. H.; Arnold, E. Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. *J. Mol. Biol.* **1996**, *264*(5), 1085-1100.
- 449 Ding, J.; Das, K.; Tantillo, C.; Zhang, W.; Clark, A. D. Jr.; Jessen, S.; Lu, X.; Hsiou, Y.; Jacobo-Molina, A.; Andries, K.; *et al.* Structure of HIV-1 reverse transcriptase in a complex with the non-nucleoside inhibitor alpha-APA R 95845 at 2.8 Å resolution. *Structure.* **1995**, *3*(4), 365-379.
- 450 Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; *et al.* High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat. Struct. Biol.* **1995**, *2*(4), 293-302.
- 451 Ren, J.; Milton, J.; Weaver, K. L.; Short, S. A.; Stuart, D. I.; Stammers, D. K. Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. *Structure.* **2000**, *8*(10), 1089-1094.
- 452 Hsiou, Y.; Das, K.; Ding, J.; Clark, A. D. Jr.; Kleim, J. P.; Rosner, M.; Winkler, I.; Riess, G.; Hughes, S. H.; Arnold, E. Structures of Tyr188Leu mutant and wild-type HIV-1 reverse transcriptase complexed with the non-nucleoside inhibitor HBY 097: inhibitor flexibility is a useful design feature for reducing drug resistance. *J. Mol. Biol.* **1998**, *284*(2), 313-323.
- 453 Gussio, R.; Pattabiraman, N.; Zaharevitz, D. W.; Kellogg, G. E.; Topol, I. A.; Rice, W. G.; Schaeffer, C. A.; Erickson, J. W.; Burt, S. K. All-atom models for the non-nucleoside binding site of HIV-1 reverse transcriptase complexed with inhibitors: A 3D QSAR approach. *J. Med. Chem.* **1996**, *39*, 1645-1650.
- 454 Hopkins, A. L.; Ren, J. S.; Esnouf, R. M.; Willcox, B. E.; Jones, E. Y.; Ross, C.; Miyasaka, T.; Walker, R. T.; Tanaka, H.; Stammers, D. K.; Stuart, D. I. Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors. *J. Med. Chem.* **1996**, *39*, 1589-1600.
- 455 Ren, J.; Esnouf, R. M.; Hopkins, A. L.; Jones, E. Y.; Kirby, I.; Keeling, J.; Ross, C. K.; Larder, B. A.; Stuart, D. I.; Stammers, D. K. 3'-Azido-3'-deoxythymidine drug resistance mutations in HIV-1 reverse transcriptase can induce long range conformational changes. *PNAS.* **1998**, *95*(16), 9518-9523.
- 456 Ren, J.; Esnouf, R. M.; Hopkins, A. L.; Warren, J.; Balzarini, J.; Stuart, D. I.; Stammers, D. K. Crystal structures of HIV-1 reverse transcriptase in complex with carboxanilide derivatives. *Biochemistry.* **1998**, *37*(41), 14394-14403.
- 457 Lane, D. M. HyperStat. Online Textbook. Accesible a través de <http://davidmlane.com/hyperstat/>. En línea (30/7/06).
- 458 Pearlman, D. A.; Charifson, P. S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem.* **2001**, *44*(21), 3417-3423.
- 459 Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions *J. Chem. Theory Comput.* **2006**, *2*(1), 128-139.
- 460 Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*(16), 1999-2012.

- 461 Maloney, P.R.; Parks, D. J., Haffner, C. D., Fivush, A. M.; Chandra, G.; Plunket, K. D.; Creech, K. L.; Moore, L. B.; Wilson, J. G.; Lewis, M. C.; Jones, S. A. and Willson, T. M. Identification of a chemical tool for the Orphan Nuclear Receptor FXR. *J. Med. Chem.* **2000**, *43*, 2971-2974.
- 462 Waldman, M.; Li, H.; Hassan, M. Novel algorithms for the optimization of molecular diversity of combinatorial libraries. *J. Mol. Graph. Model.* **2000**, *18*, 412-426, 533-536.
- 463 Higgs, R. E.; Bemis, K. G.; Watson, I. A.; Wikel, J. H. Experimental Designs for Selecting Molecules from Large Chemical Databases. *J. Chem. Inf. Comput. Sci.* **1997**, *37*(5), 861-870.
- 464 StatGraphics. StatPoint, Inc. Northern Virginia. USA.
- 465 Everitt, B. S.; Landau, S.; Leese, M. *Cluster Analysis*. John Wiley & Sons Inc: 2001; pp 128-129.
- 466 De Clercq, E.; Anti-HIV chemotherapy: current state of the art. *Med. Chem. Res.* **2004**, *13*(6/7), 439-478.
- 467 Markovic, I.; Clouse, K. A. Recent advances in understanding the molecular mechanisms of HIV-1 entry and fusion: revisiting current targets and considering new options for therapeutic intervention. *Curr. HIV Res.* **2004**, *2*, 223-234.
- 468 Lapham, C. K.; Ouyang, J.; Chandrasekhar, B.; Nguyen, N. Y.; Dimitrov, D. S.; Golding, H. Evidence for cell-surface association between fusin and the CD4-gp120 complex in human cell lines. *Science.* **1996**, *274*, 602-605.
- 469 Alkhatib, G.; Combadiere, C.; Broder, C. C.; Feng, Y.; Kennedy, P. E.; Murphy, P. M.; Berger, E. A. CC CKR5: a RANTES, MIP-1, MIP-1 receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science.* **1996**, *272*, 1955-1958.
- 470 Shaheen, F.; Collman, R. G. Co-receptor antagonist as HIV-1 entry inhibitors. *Curr. Opin. Infect. Dis.* **2004**, *17*(1), 7-16.
- 471 Kazmierski, W. M.; Peckman, J. P.; Duan, M.; Kenakin, T. P.; Jenkinson, S.; Gudmundsson, K. S.; Piscitelli, S. C.; Feldman, P. L. Recent Progress in the Discovery of New CCR5 and CXCR4 Chemokine Receptor Antagonists as Inhibitors of HIV-1 Entry. Part 2. *Curr. Med. Chem. – Anti Infective Agents.* **2005**, *4*, 133-152.
- 472 Zou, Y.R.; Kottmann, A. H.; Kuroda, M.; Taniuchi, I.; Littman, D. R. Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development. *Nature.* **1998**, *393*, 595-599.
- 473 De Clercq, E. Inhibition of HIV Infection by Bicyclams, Highly Potent and Specific CXCR4 Antagonists. *Mol. Pharmacol.* **2000**, *57*, 833-839.
- 474 Esté, J. A.; Cabrera, C.; De Clercq, E.; Struyf, S.; Van Damme, J.; Bridger, G.; Skerlj, R. T.; Abrams, M. J.; Henson, G.; Gutierrez, A.; Clotet, B.; Schols, D. Activity of Different Bicyclam Derivatives against Human Immunodeficiency Virus Depends on Their Interaction with the CXCR4 Chemokine Receptor. *Mol. Pharmacol.* **1999**, *55*(1), 67-73.
- 475 Egberink, H. F.; De Clercq, E.; Van Vliet, A. L. W.; Balzarini, J.; Bridger, G. J.; Henson, G.; Horzinek, M. C.; Schols, D. Bicyclams, Selective Antagonists of the Human Chemokine Receptor CXCR4, Potently Inhibit Feline Immunodeficiency Virus Replication. *J. Virol.* **1999**, *73*(8), 6346-6352.
- 476 De Clercq, E.; Yamamoto, N.; Pauwels, R.; Baba, M.; Schols, D.; Nakashima, H.; Balzarini, J.; Debyser, Z.; Murrer, B. A.; Schwartz, D. Potent and selective inhibition of human immunodeficiency virus (HIV)-1 and HIV-2 replication by a class of bicyclams interacting with a viral uncoating event. *PNAS.* **1992**, *89*(12), 5286-5290.

- 477 Schols, D.; Struyf, S.; Van Damme, J.; Esté J, A.; Henson, G.; De Clercq, E. Inhibition of T-tropic HIV Strains by Selective Antagonization of the Chemokine Receptor CXCR4. *J. Exp. Med.* **1997**, *186*(8), 1383-1388.
- 478 Hatse, S.; Princen, K.; De Clercq, E.; Rosenkilde, M. M.; Schwartz, T. W.; Hernandez-Abad, P. E.; Skerlj, R. T.; Bridger, G. J.; Schols, D. AMD3465, a monomacrocyclic CXCR4 antagonist and potent HIV entry inhibitor. *Biochem. Pharmacol.* **2005**, *70*(5), 752-761.
- 479 Princen, K.; Hatse, S.; Vermeire, K.; Aquaro, S.; De Clercq, E.; Gerlach, L. O.; Rosenkilde, M. M.; Schwartz, T. W.; Skerlj, R.; Bridger, G.; Schols, D. Inhibition of Human Immunodeficiency Virus Replication by a Dual CCR5/CXCR4 Antagonist. *J. Virol.* **2004**, *78*(23), 12996-13006.
- 480 Bridger, G. J.; Skerlj, R. T.; Padmanabhan, S.; Martellucci, S. A.; Henson, G. W.; Struyf, S.; Witvrouw, M.; Schols, D.; De Clercq, E. Synthesis and structure-activity relationships of phenylenebis(methylene)-linked bis-azamacrocycles that inhibit HIV-1 and HIV-2 replication by antagonism of the chemokine receptor CXCR4. *J. Med. Chem.* **1999**, *42*(19), 3971-3981.
- 481 Bridger, G.; Skerlj, R.; Kaller, A. L.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y.; Schols, D.; Smith C. D.; Di Fluri, R.M. Chemokine receptor binding heterocyclic compounds. WO0222600. **2002**.
- 482 Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y.; Schols, D.; Smith C. D.; Di Fluri, R.M. Chemokine receptor binding heterocyclic compounds. WO0222599, **2002**.
- 483 Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y.; Schols, D.; Smith C. D.; Di Fluri, R.M. Chemokine receptor binding heterocyclic compounds. WO0234745, **2002**.
- 484 Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y.; Schols, D. Chemokine receptor binding heterocyclic compounds. WO0056729, **2003**.
- 485 Bridger, G.; McEachern, E. J.; Skerlj, R.; Schols, D.; Baird I.; Kaller, A.; Curtis, H.; Zhu, Y.; Gang, C.; Krystyna, S. CXCR4 Chemokine receptor binding compounds. WO2004091518, **2004**.
- 486 Ichiyama, K.; Yokohama-Kumakura, S.; Tanaka, Y.; Tanaka, R.; Hirose, K.; Bannai, K.; Edamatsu, T.; Yanaka, M.; Niitani, Y.; Miyako-Kurosaki, N.; Takaku, H.; Koyanagi, Y.; Yamamoto, N. A duodenally absorbable CXC chemokine receptor 4 antagonist, KRH-1636, exhibits a potent and selective anti-HIV-1 activity. *PNAS.* **2003**, *100*(7), 4185-4190.
- 487 Yamazaki, T.; Saitou, A.; Ono, M.; Yokohama, S.; Bannai, K.; Hirose, K.; Yanaka, M. Novel nitrogenous compound and use thereof. WO03029218, **2003**.
- 488 Yamazaki, T.; Kikumoto, S.; Ono, M.; Saitou, A.; Takahashi, H.; Kumakura, S.; Hirose, K.; Yanaka, M.; Takemura, Y.; Suzuki, S.; Matsui, R. Amine compounds and use thereof. WO2004024697, **2004**.
- 489 Masuda, M.; Nakashima, H.; Ueda, T.; Naba, H.; Ikoma, R.; Otaka, A.; Terakawa, Y.; Tamamura, H.; Ibuka, T.; Murakami, T.; *et al.* A novel anti-HIV synthetic peptide, T-22 ([Tyr5,12,Lys7]-polyphemusin II). *Biochem. Biophys. Res. Comm.* **1992**, *189*(2), 845-850.
- 490 Arakaki, R.; Tamamura, H.; Premanathan, M.; Kanbara, K.; Ramanan, S.; Mochizuki, K.; Baba, M.; Fujii, N.; Nakashima, H. T134, a small molecule CXCR4 inhibitor, has no cross-drug resistance with AMD3100, a CXCR4 antagonist with a different structure. *J. Virol.* **1999**, *73*(2), 1719-1723.

- 491 Tamamura, H.; Omagari, A.; Hiramatsu, K.; Gotoh, K.; Kamamoto, T.; Xu, Y.; Kodama, E.; Matsuoka, M.; Hattori, T.; Yamamoto, N.; Nakashima, H.; Otaka, A.; Fujii, N. Development of specific CXCR4 inhibitors possessing high selectivity indexes as well as complete stability in serum based on an anti-HIV peptide T140. *Bioorg. Med. Chem. Lett.* **2001**, *11*(14), 1897-1902.
- 492 Summer-Smith, M.; Zheng, Y.; Zhang, Y. P.; Twist, E. M.; Climie, S. C. Antitherpetic activities of N-alpha-acetyl-nona-D-arginine amide acetate. *Drugs Exp. Clin. Res.* **1995**, *21*(1), 1-6.
- 493 Kazmierski, W. M.; Aquino, C. J.; Bifulco, N.; Boros, E. E.; Chauder, B. A.; Chong, P. Y.; Duan, M.; Deanada, F. Jr.; Koble, C. S.; Mclean, E. W.; Peckham, J. P.; Perkins, A. C.; Thompson, J. B.; Vanderwall, D. CCR5 Antagonists as therapeutic agents. WO2004054974, **2004**.
- 494 Duan, M.; Kazmierski, W. M.; Aquino, C. J. Cyclohexyl compounds as CCR5 antagonists. WO2004054581, **2004**.
- 495 Peckham, J. P.; Aquino, C. J.; Kazmierski, W. M. Cyclopropyl compounds as CCR5 antagonists. WO2004055010, **2004**.
- 496 Aquino, C. J.; Chong, P. Y.; Duan, M.; Kazmierski, W. M. Heterocyclic compounds as CCR5 antagonists. WO2004055011, **2004**.
- 497 Youngman, M.; Kazmierski, W. M.; Yang, H.; Aquino, C. J. Indane compounds as CCR5 antagonists. WO2004055012, **2004**.
- 498 Yang, H.; Kazmierski, W. M.; Aquino, C. J. Pyrrolidine and Azetidine compounds as CCR5 antagonists. WO2004055016, **2004**.
- 499 Maeda, K.; Yoshimura, K.; Shibayama, S.; Habashita, H.; Tada, H.; Sagawa, K.; Mikayawa, T.; Aoki, M.; Fukushima, D.; Mitsuya, H. Novel Low Molecular Weight Spirodiketopiperazine Derivatives Potently Inhibit R5 HIV-1 Infection through Their Antagonistic Effects on CCR5. *J. Biol. Chem.* **2001**, *276*(37), 35194-35200.
- 500 Shibayama, S.; Sagawa, K.; Watanabe, N.; Takeda, K.; Tada, H.; Fukushima, D. Antagonist and agonist binding to strong binding site of chemokine receptor. WO2004054616, **2004**.
- 501 Takaoka, Y.; Okamoto, M.; Genba, Y. Novel crystals of triazaspiro[5.5]undecane derivative. WO2004026874, **2004**.
- 502 Takaoka, Y.; Nishizawa, R.; Shibayama, S.; Sagawa, K.; Matsuo, M. Y. WO2002074770, **2002**.
- 503 Imawaka, H.; Shibayama, S.; Takaoka, Y. WO2003035074, **2003**.
- 504 Palani, A.; Shapiro, S.; Clader, J. W.; Greenlee, W. J.; Blythin, D.; Cox, K.; Wagner, N. E.; Strizki, J.; Baroudy, B. M.; Dan N. Biological evaluation and interconversion studies of rotamers of SCH 351125, an orally bioavailable CCR5 antagonist. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 705-708.
- 505 Tsamis, F.; Gavrilov, S.; Kajumo, F.; Seibert, C.; Kuhmann, S.; Ketas, T.; Trkola, A.; Palani, A.; Clader, J. W.; Tagat, J. R.; McCombie, S.; Baroudy, B.; Moore, J. P.; Sakmar, T. P.; Dragic, T. J. Analysis of the mechanism by which the small-molecule CCR5 antagonists SCH-351125 and SCH-350581 inhibit human immunodeficiency virus type 1 entry. *J. Virol.* **2003**, *77*(9), 5201-5208.
- 506 Billick, E.; Seibert, C.; Pugach, P.; Ketas, T.; Trkola, A.; Endres, M. J.; Murgolo, N. J.; Coates, E.; Reyes, G. R.; Baroudy, B. M.; Sakmar, T. P.; Moore, J. P.; Kuhmann, S. E. The Differential Sensitivity of Human and Rhesus Macaque CCR5 to Small-Molecule Inhibitors of Human Immunodeficiency Virus Type 1 Entry Is Explained by a Single Amino Acid Difference and Suggests a Mechanism of Action for These Inhibitors. *J. Virol.* **2004**, *78*, 4134-4144.

- 507 Aramaki, Y.; Seto, M.; Okawa, T.; Oda, T.; Kanzaki, N.; Shiraishi, M. Synthesis of 1-benzothiepine and 1-benzazepine derivatives as orally active CCR5 antagonists. *Chem. Pharm. Bull.* **2004**, *52*, 254-258.
- 508 Seto, M.; Aramaki, Y.; Okawa, T.; Miyamoto, N.; Aikawa, K.; Kanzaki, N.; Shiraishi, M. Orally active antagonists as Anti-HIV-1 Agents: Synthesis and Biological Activity of 1-Benzothiepine 1,1-Dioxide and 1-Benzazepine derivatives containing a tertiary Amine Moiety. *Chem. Pharm. Bull.* **2004**, *52*, 577-590.
- 509 Imamura, S.; Ishihara, Y.; Hattori, T.; Kurasawa, O.; Matsushita, Y.; Sugihara, Y.; Kanzaki, N.; Lizawa, Y.; Baba, M.; Hashiguchi, S. CCR5 antagonists as Anti-HIV-1 agents. 1. Synthesis and Biological Evaluation of 5-oxopyrrolidine-3-carboxamide Derivatives. *Chem. Pharm. Bull.* **2004**, *52*, 63-73.
- 510 Imamura, S.; Kurasawa, O.; Nara, Y.; Ichikawa, T.; Nishikawa, Y.; Iida, T.; Hashiguchi, S.; Kanzaki, N.; Lizawa, Y.; Baba, M.; Sugihara, Y. CCR5 antagonists as anti-HIV-1 agents. Part 2: Synthesis and biological evaluation of N-[3-(4-benzylpiperidin-1-yl)propyl]-N,N'-diphenylureas. *Bioorg. Med. Chem.* **2004**, *12*, 2295-2306.
- 511 Willouhby, C. W.; Rosauer, K. G.; Hale, J. J.; Budhu, R. J.; Mills, S. G.; Chapman, K. T.; MacCoss, M.; Malkowitz, L.; Springer, M. S.; Gould, S. L.; DeMartino, J. A.; Siciliano, S. J.; Cascieri, M. A.; Carella, A.; Catver, G.; Colmes, K.; Schlieff, W. A.; Danzeisen, R.; Hazuda, D.; Kessler, J.; Lineberger, J.; Miller, M.; Emini, E. A. 1,3,4 Trisubstituted pyrrolidine CCR5 receptor antagonists bearing 4-aminoheterocycle substituted piperidine side chains. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 427-431.
- 512 Rusconi, S.; Scozzafava, A.; Mastrolorenzo, A.; Supuran, T. C.; New Advances in HIV Entry Inhibitors Development. *Curr. Drug Targets Infect. Disord.* **2004**, *4(4)*, 339-355.
- 513 Perros, M.; Price, D. A.; Stammen, B. L. C.; Wood, A. Tropane derivatives useful in therapy. NZ521477, **2003**.
- 514 Basford, P. A.; Stephenson, P. T.; Taylor, S. C. J.; Wood, A. Tropane derivatives as CCR5 modulators. WO03084954, **2003**.
- 515 Armour, D. R.; Price, D. A.; Stammen, B. L. C.; Wood, A.; Perros, M.; Edwards, M. P. Piperidine as CCR5 modulator. JP2004099618, **2004**.
- 516 Cumming, J.; Winter, J. N-4-piperidinyl compounds as CCR4 modulators. WO2004018425, **2004**.
- 517 Burrows, J.; Cumming, J.; Tucker, H. N-4-piperidinyl compounds as CCR5 modulators. US2004110952, **2004**.
- 518 Cumming, J. Novel piperidine derivatives as modulators of chemokine receptors. WO03042178, **2003**.
- 519 Cumming, J. Piperidine or 8-aza-bicyclo[3.2.1]oct-3-yl derivatives useful as modulators of chemokine receptor activity (especially CCR5). WO033080574, **2003**.
- 520 Cumming, J.; Tucker, H. Piperidine derivatives and their use as modulators of chemokine receptor activity (especially CCR5). WO03042177, **2003**.
- 521 Doranz, B. J.; Grovit-Ferbas, K.; Sharron, M. P.; Mao, S. H.; Gotees, M. B.; Daar, E. S.; Doms, R. W.; O'Brien, W. A. A small-molecule inhibitor directed against the chemokine receptor CXCR4 prevents its use as an HIV-1 coreceptor. *J. Exp. Med.* **1997**, *186*, 1395-1400.
- 522 Labrosse, B.; Brelot, A.; Heveker, N.; Sol, N.; Schols, D.; De Clercq, E.; Alizon, M. Determinants for Sensitivity of Human Immunodeficiency Virus Coreceptor CXCR4 to the Bicyclam AMD3100. *J. Virol.* **1998**, *72*, 6381-6388.

- 523 Hatse, S.; Princen, K.; Gerlach, L.; Bridger, G.; Henson, G.; De Clerq, E.; Schwartz, T.; Schols, D. Mutation of Asp171 and Asp262 of the Chemokine Receptor CXCR4 Impairs Its Coreceptor Function for Human Immunodeficiency Virus-1 Entry and Abrogates the Antagonistic Activity of AMD3100. *Mol. Pharmacol.* **2001**, *60*, 164-173.
- 524 Gerlach, L.; Skerlj, R.; Bridger, G.; Schwartz, T. Molecular Interactions of Cyclam and Bicyclam Non-peptide Antagonists with the CXCR4 Chemokine Receptor. *J. Biol. Chem.* **2001**, *276*(17), 14153-14160.
- 525 Rosenkilde, M. M.; Gerlach, L. O.; Jakobsen, J. S.; Skerlj, R. T.; Bridger, G. J.; Schwartz, T. W. Molecular Mechanism of AMD3100 Antagonism in the CXCR4 Receptor. *J. Biol. Chem.* **2004**, *279*, 3033-3041.
- 526 Trent, J. O.; Wang, Z.; Murray, J. L.; Shao, W.; Tamamura, H.; Fujii, N.; Peiper, S. C. Lipid Bilayer Simulations of CXCR4 with Inverse Agonist and Weak Partial Agonist. *J. Biol. Chem.* **2003**, *36*, 47136-47144.
- 527 Liang, X.; Parkinson, J. A.; Weishäupl, M.; Gould, R. O.; Paisey, S. J.; Park, H.; Hunter, T. M.; Blindauer, C. A.; Parsons, S.; Sadler, P. J. Structure and Dynamics of Metallomacrocycles: Recognition of Zinc Xylyl-Bicyclam by an HIV Coreceptor. *J. Am. Chem. Soc.* **2002**, *124*, 9105-9112.
- 528 Baba, M.; Nishimura, O.; Kanzaki, N.; Okamoto, M.; Sawada, H.; Iizawa, Y.; Shirashi, M.; Aramki, Y.; Okonogi, K.; Ogawa, Y.; Meguro, K.; Fujino, M. A small-molecule, nonpeptide CCR5 antagonist with highly potent and selective anti-HIV-1 activity. *PNAS.* **1999**, *96*, 5698-5703.
- 529 Dragic, T.; Trkola, A.; Thompson, D. A.; Cormier, E. G.; Kajumo, F. A.; Maxwell, E.; Lin, S. W.; Ying, W.; Smith, S. O.; Sakmar, T. P.; Moore, J. P.; A binding pocket for a small molecule inhibitor of HIV-1 entry within the transmembrane helices of CCR5. *PNAS.* **2000**, *97*, 5639-5644.
- 530 Tsamis, F.; Gavrillov, S.; Kajumo, F.; Seibert, C.; Kuhmann, S.; Ketas, T.; Trkola, A.; Palani, A.; Clader, J. W.; Tagat, J. R.; McCombie, S.; Baroudy, B.; Moore, J. P.; Sakmar, T. P.; Dragic, T. Analysis of the Mechanism by Which the Small-Molecule CCR5 Antagonists SCH-351125 and SCH-350581 Inhibit Human Immunodeficiency Virus Type 1 Entry. *J. Virol.* **2003**, *77*(9), 5201-5208.
- 531 Seibert, C.; Ying, W.; Gavrillov, S.; Tsamis, F.; Kuhmann, S. E.; Palani, A.; Tagat, J. R.; Clader, J. W.; McCombie, S. W.; Baroudy, B. M.; Smith, S. O.; Dragic, T.; Moore, J. P.; Sakmar, T. P. Interaction of small molecule inhibitors of HIV-1 entry with CCR5. *Virology.* **2006**, *349*(1), 41-54.
- 532 Paterlini, M. G. Structure Modeling of the Chemokine Receptor CCR5: Implications for Ligand Binding and Selectivity. *Biophys. J.* **2002**, *83*, 3012-3031.
- 533 Teller, D. C.; Okada, T.; Behnke, C. A.; Palczewski, K.; Stenkamp, R. E. Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). *Biochemistry.* **2001**, *40*(26), 7761-7772.
- 534 Ballesteros, J. A.; Shi, L.; Javitch, J. A. Structural Mimicry in G Protein-Coupled Receptors: Implications of the High-Resolution Structure of Rhodopsin for Structure-Function Analysis of Rhodopsin-Like Receptors. *Mol. Pharmacol.* **2001**, *60*, 1-19.
- 535 Bruccoleri, R. E. Application of Systematic Conformational Search to Protein Modeling. *Molecular Simulations.* **1993**, *10*, 151-174. Accesible a través de <http://www.CONGEN.com>. En línea (15/2/06).

- 
- 536 Bruccolieri, R. E. Ab initio Loop Modeling and its Application to Homology Modeling. En *Proteins Structure Prediction. Methods and Protocols*. Methods in Molecular Biology; Webster, D.M., Ed.; Humana Press: UK: 2000, pp 143.
- 537 Laskowski, R. A.; McArthur, M. W.; Moss, D. S.; and Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **1993**, *26*, 283-291. Accesible a través de <http://www.biochem.ucl.ac.uk/%7Eroman/procheck/procheck.html>. En línea (15/2/06).
- 538 Sroczynski, D.; Grzejdzia, A. Physical Image vs. Structure Relation. Part 3[1]. Basic Properties and Protonation Mechanism of Some Tetraaza Macrocyclic Ligands. *J. Incl. Phenom. Macro.* **1999**, *35*, 251-260.
- 539 McDonald, I. K.; Thornton, J. M. Satisfying Hydrogen Bonding Potential in Proteins. *J. Mol. Biol.* **1994**, *238*, 777-793. Accesible a través de <http://www.biochem.ucl.ac.uk/bsm/hbplus/home.html>. En línea (11/4/06).
- 540 Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins.* **2003**, *50*, 5-25.
- 541 Pettersson, S. Comunicación personal. 2006.

