



**Universitat Ramon Llull**

## **TESIS DOCTORAL**

Título	<b>Diseño, selección y síntesis de nuevos inhibidores de entrada del VIH.</b>
Realizada por	Sofia Henriette Pettersson Salom
en el Centro	Escola Tècnica Superior IQS
y en el Departamento	Química Orgánica
Dirigida por	Dr. Jordi Teixidó i Closa



## **CAPÍTULO 4. FARMACÓFOROS**



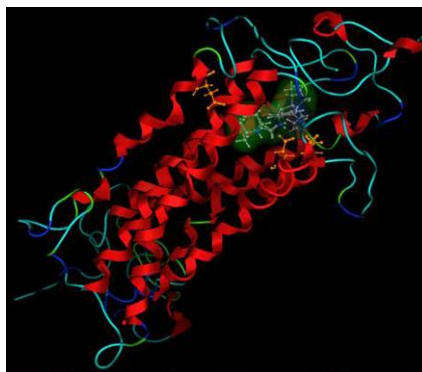
## **4.1. Introducción a los farmacóforos.**

### **4.1.1. Cribado virtual.**

El aumento de la presión económica sobre la industria farmacéutica para desarrollar nuevos fármacos de manera más rápida y eficiente ha llevado al desarrollo de nuevas técnicas en el proceso de descubrimiento de nuevos *leads* y optimización de los candidatos a fármacos. Los avances producidos en el campo de la síntesis orgánica con el desarrollo de la química combinatoria han permitido sintetizar grandes quimiotecas de compuestos<sup>[1]</sup> y los métodos *high-throughput screening* (HTS) han reducido el tiempo necesario para determinar la actividad biológica de las moléculas.<sup>[2-4]</sup> La posibilidad de evaluar todos los compuestos sintetizables frente a todas las proteínas disponibles es todavía un objetivo inalcanzable, así como el bajo rendimiento de *leads* identificados en los procesos HTS. Por lo tanto se necesita disponer de modelos que permitan seleccionar *in silico* posibles fármacos dentro de bases de datos de compuestos (sintetizados o propuestos). Estas técnicas se conocen con el nombre de *virtual screening*<sup>[5]</sup> o cribado virtual.<sup>[6]</sup> El objetivo principal del cribado virtual es reducir el espacio químico representado por las moléculas orgánicas pequeñas a un número manejable de compuestos con la mayor posibilidad de convertirse en candidatos a fármacos para la diana de interés.

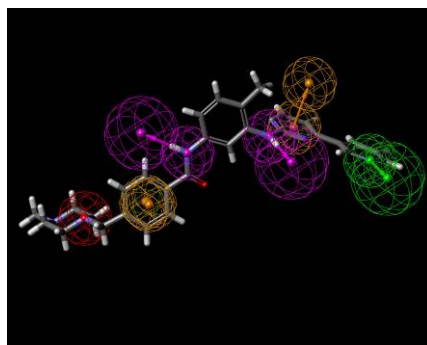
Las técnicas de cribado virtual se pueden clasificar en dos grupos según la información disponible sobre la estructura del receptor y de los ligandos del receptor:<sup>[7]</sup>

- Cribado virtual basado en la estructura del receptor (*structure-based virtual screening*), utiliza la información de la estructura de la proteína resuelta por rayos-X o RMN para seleccionar aquellas moléculas que presentan la mejor disposición en el sitio de unión del receptor mediante técnicas de *docking*, o diseñar moléculas a partir de la estructura del receptor mediante técnicas de diseño *de novo*. Este tipo de cribado virtual ha sido realizado por la Dra. Pérez-Nueno en el Laboratorio de Diseño Molecular del GEM.<sup>[8]</sup>



**Figura 4.1.** Docking de una molécula sobre el receptor CXCR4.<sup>[9]</sup>

- Cribado virtual basado en los ligandos (*ligand-based virtual screening*), utiliza la información derivada de las similitudes estructurales de los ligandos de una determinada proteína para seleccionar moléculas que presenten características estructurales semejantes mediante modelos farmacofóricos, análisis de similitud de compuestos de una base de datos respecto a un compuesto de referencia, o QSAR. Este tipo de cribado virtual se realiza en el presente trabajo.



**Figura 4.2.** Modelo farmacofórico (figura de [www.accelrys.com/products/discovery-studio/pharmacophores](http://www.accelrys.com/products/discovery-studio/pharmacophores)).

A principios del s. XX, Ehrlich introduce el concepto de farmacóforo como las características de una molécula (*phoros*) que son esenciales para su actividad biológica (*pharmacon*).<sup>[10]</sup> La definición actual fue desarrollada por Gund en 1977 y describe un farmacóforo como un conjunto de características estructurales en una molécula que es reconocido en el sitio de unión del receptor y es responsable de la actividad biológica de la molécula.<sup>[11]</sup>

Los modelos farmacofóricos, o farmacóforos describen la disposición espacial de una serie de características claves para la actividad de compuestos frente a una diana terapéutica. Un farmacóforo no representa una molécula real o una asociación real de grupos funcionales, sino un concepto abstracto que reúne las interacciones moleculares comunes de un grupo de compuestos activos. Estos modelos proporcionan información visual y cualitativa, capturan la naturaleza de la interacción del ligando con el receptor. Generalmente las predicciones de actividad derivadas de modelos farmacofóricos son de tipo binario, activo o inactivo.

Se han desarrollado varios paquetes de *software* para el modelado de farmacóforos, aquellos de uso más extendido son DISCO (*Distance Comparisons*),<sup>[12]</sup> Catalyst,<sup>[13]</sup> GASP (*Genetic Algorithm Superimposition Program*),<sup>[14]</sup> ALLADIN<sup>[15]</sup> y MOE.<sup>[16]</sup>

### 4.1.2. Desarrollo de farmacóforos con MOE.

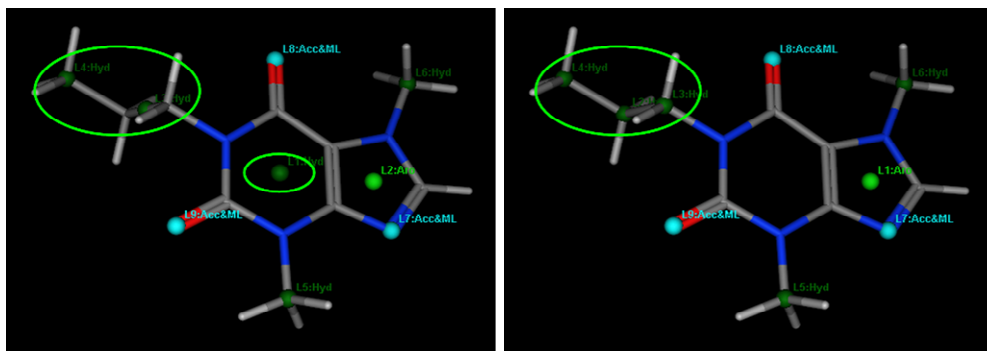
En MOE,<sup>[16]</sup> los farmacóforos están formados por la anotación de ligandos, que se define como un conjunto de puntos en el espacio que reflejan la presencia o ausencia de una o varias *features* o características farmacofóricas con un radio variable. Estas *features* pueden ser dador (Don) y aceptor (Acc) de puente de hidrógeno, presencia de cationes (Cat), aniones (Ani), centros aromáticos (Aro), zonas hidrofóbicas (Hyd) y ligandos de metales (ML); además se incluyen volúmenes tanto de exclusión, en el interior de los cuales no se puede haber ningún átomo, como de inclusión, donde sí debe posicionarse un átomo. Además se utiliza un esquema farmacofórico, que es una colección de funciones que definen el significado, apariencia y métodos de cálculo de la anotación de los ligandos. Los esquemas farmacofóricos implementados en MOE son los siguientes (Tabla 4.1):

- PCH, *Polarity-Charge-Hydrophobicity* (Figura 4.3):

Seleccionado en MOE por defecto. Incluye átomos dadores y aceptores de puente de hidrógeno, cationes, aniones, centros aromáticos y áreas hidrofóbicas que se anotan como *clusters* de átomos hidrofóbicos y centros de anillos aromáticos.

- PCH\_All (Figura 4.3):

Se diferencia del anterior en que los átomos hidrofóbicos no aromáticos se anotan individualmente.



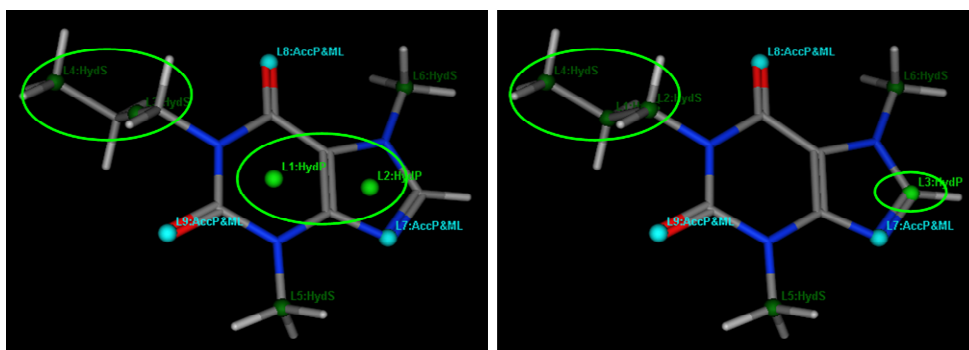
**Figura 4.3.** Esquema de anotación PCH (izquierda) y PCH\_All (derecha). Se observa la diferencia en las áreas hidrofóbicas (marcadas con elipses verdes).

- PPCH, *Planar-Polarity-Charge-Hydrophobicity* (Figura 4.4):

Diferencia entre átomos dadores y aceptores de puente de hidrógeno y áreas hidrofóbicas planares ( $sp^2$ ) y no planares ( $sp^3$ ).

- PPCH\_All (Figura 4.4):

Se diferencia del anterior en que los átomos hidrofóbicos se anotan individualmente en vez de agruparse en un área.



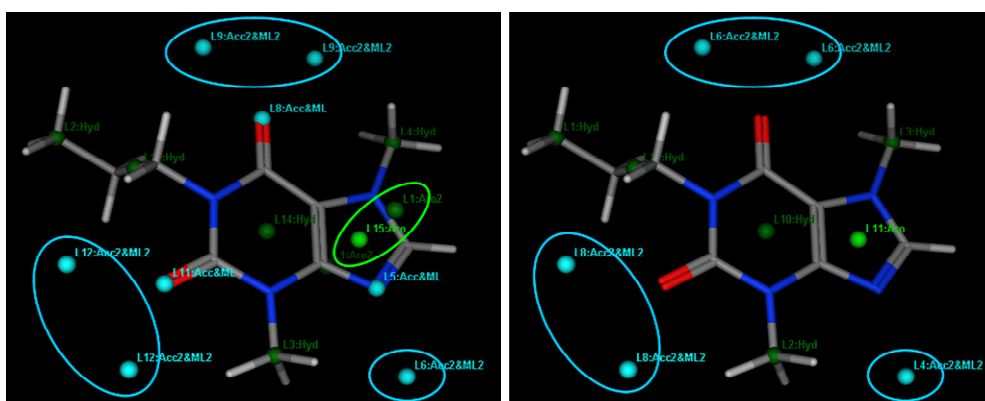
**Figura 4.4.** Esquema de anotación PPCH (izquierda) y PPCH\_All (derecha). Se observa la diferencia en las áreas hidrofóbicas (marcadas con elipses verdes).

- PCHD, *Polarity-Charge-Hydrophobicity-Direction* (Figura 4.5):

Se basa en el esquema PCH pero además genera *site points* que representan la posición hipotética de átomos complementarios en un receptor, determinados a partir de la posición de los átomos pesados en el ligando, de manera que presenta unos puntos putativos proyectados a partir de dadores y aceptores de puente de hidrógeno y centros aromáticos.

- CHD, *Charge-Hydrophobicity-Direction* (Figura 4.5):

Se basa en el esquema PCHD pero ignora los *features* definidos como *clusters* de dadores, aceptores de puente de hidrógeno y ligandos de metales.



**Figura 4.5.** Esquema de anotación PCHD (izquierda) y CHD (derecha). Se observan las características farmacofóricas basadas en los *site points* (marcadas con elipses).



**Tabla 4.1.** Anotaciones de los esquemas farmacofóricos definidos en MOE, en las características farmacofóricas la terminación P se refiere planar, la terminación S a no planar y la terminación 2 a punto proyectado.

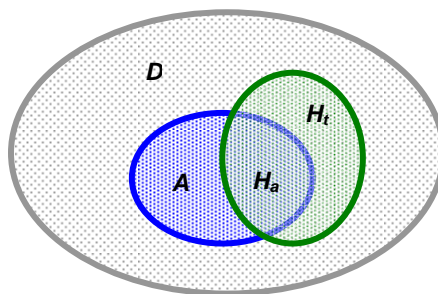
	PCH	PCH_AII	PPCH	PPCH_AII	CHD	PCHD
Don	X	X				X
DonP			X	X		
DonS			X	X		
Don2					X	X
Acc	X	X				
AccP			X	X		
AccS			X	X		
Acc2					X	X
Cat	X	X	X	X	X	X
Ani	X	X	X	X	X	X
Aro	X	X			X	X
Aro2						X
Hyd	X	X			X	X
HydP			X	X		
HydS			X	X		
ML	X	X	X	X		X
ML2					X	X

## 4.2. Parámetros de análisis de los farmacóforos.

A partir del modelo farmacofórico generado se realiza un análisis retrospectivo sobre la base de datos multiconformacional (base de datos de conformaciones de cada uno de los compuestos considerados) de activos e inactivos, donde se emparejan todos los puntos de anotación de los compuestos de la base de datos con cada una de las características farmacofóricas definidas en el modelo. El resultado obtenido es una lista de compuestos considerados como activos, *hit list*, que se ordena según las distancias cuadráticas promedio, expresado como RMSD, entre cada característica farmacofórica del modelo y el punto correspondiente en la conformación evaluada.

Las métricas que se utilizan para comparar los diferentes farmacóforos obtenidos se presentan a continuación:<sup>[17,18]</sup>

- $D$ , número de compuestos de la base de datos.
- $A$ , número de compuestos activos de la base de datos.
- $H_t$ , número de compuestos en el *hit list* obtenido al cribar una base de datos con el modelo farmacofórico en estudio.
- $H_a$ , número de compuestos activos en el *hit list* al cribar una base de datos con el modelo farmacofórico en estudio.



**Figura 4.6.** Representación del conjunto de compuestos de la base de datos,  $D$ , los activos,  $A$ , el *hit list* obtenido con el farmacóforo,  $H_t$ , y los compuestos activos del *hit list*,  $H_a$ .

- Falsos negativos, definidos según la Ecuación 4.1.

$$falsos- = A - H_a \quad [4.1]$$

- Falsos positivos, definidos según la Ecuación 4.2.

$$falsos+ = H_t - H_a \quad [4.2]$$

El mejor hit list se obtiene cuando se da una superposición perfecta entre el hit list y los activos de la base de datos, ocurre cuando se satisfacen las condiciones  $H_a=H_t$  y  $H_a=A$ .

- %Y, rendimiento de activos, representa la relación entre el número de compuestos activos del *hit list* obtenido al cribar una base de datos con un farmacóforo y el número de compuestos totales en dicho *hit list*, y se define mediante la Ecuación 4.3. Cuanto mayor es su valor más selectivo es el farmacóforo sobre el dominio de la base de datos.

$$\%Y = \frac{H_a}{H_t} \cdot 100 \quad [4.3]$$

- %A, ratio de activos, se expresa como la relación entre el número de compuestos activos presentes en el *hit list* obtenido al cribar una base de datos con el farmacóforo y el número de activos reales de dicha base de datos, y se puede expresar mediante la Ecuación 4.4.

$$\%A = \frac{H_a}{A} \cdot 100 \quad [4.4]$$

Los parámetros %Y y %A no son totalmente independientes entre ellos, y generalmente, al intentar maximizar uno de ellos se compromete el otro. Sin embargo, para obtener un buen modelo farmacofórico se deben maximizar ambos parámetros simultáneamente.

- *ef*, factor de enriquecimiento (*enrichment factor*), indica cuántas veces más enriquecida en activos está el *hit list* obtenido al cribar una base de datos con un farmacóforo respecto a la base de datos original, y se calcula mediante la Ecuación 4.5. Un valor superior a 1 indica que la selección de compuestos mediante el modelo evaluado es mejor que una selección al azar. Por ejemplo, un *ef* de 2 indica que es dos veces más probable seleccionar al azar un compuesto activo en el *hit list* que de la base de datos original.

$$ef = \frac{H_a/H_t}{A/D} \quad [4.5]$$

Un dato a tener en cuenta es que en el estudio realizado en el presente trabajo, se dispondrá de una base de datos compuesta por 151 activos y 1613 inactivos, el máximo *ef* que se puede obtener en los modelos farmacofóricos generados es de 10.68, y ocurre cuando  $H_a=H_t=A$ .

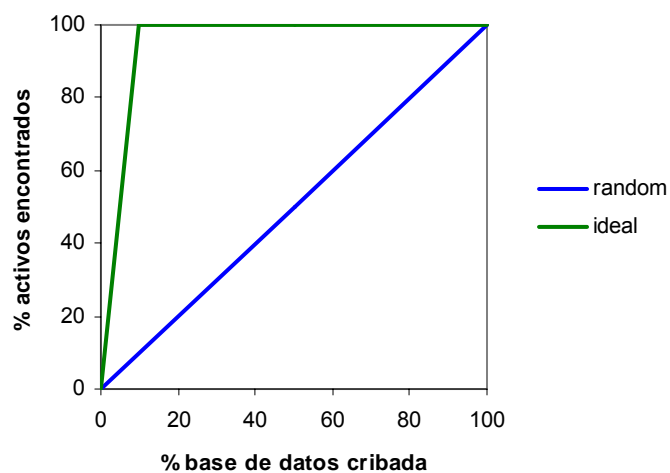
- *GH*, bondad (*goodness*) del *hit list*, representa el compromiso entre el máximo rendimiento y el máximo porcentaje de activos encontrados, y se expresa mediante la Ecuación 4.6.<sup>[19]</sup> Cuanto más cercano a 1 es su valor, mejor es el *hit list* obtenido al cribar la base de datos con el farmacóforo.

$$GH = \frac{H_a(3A + H_t)}{4H_tA} \left(1 - \frac{H_t - H_a}{D - A}\right) \quad [4.6]$$

- $GH_r$ , componente aleatoria de la bondad del *hit list* obtenido, su valor se calcula a partir de la Ecuación 4.7.<sup>[20]</sup>

$$GH_r = \left(\frac{3}{4} \frac{A}{D} + \frac{1}{4} \frac{H_t}{D}\right) \left(1 - \frac{H_t}{D}\right) \quad [4.7]$$

Se representan también las curvas de enriquecimiento, donde se muestra el porcentaje de activos seleccionados mediante el farmacóforo respecto al porcentaje de compuestos cribados de la base de datos. En la Figura 4.7 se observa el gráfico correspondiente a una selección al azar y a un modelo ideal donde al cribar la base de datos primero se seleccionan todos los activos.



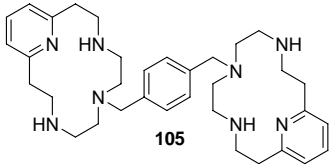
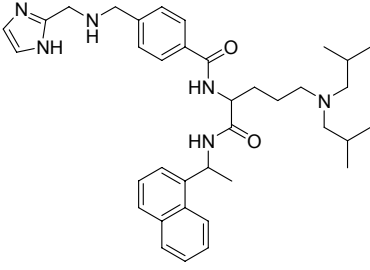
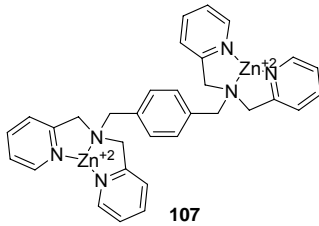
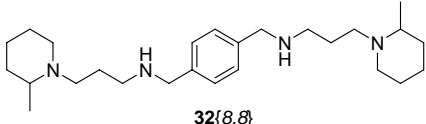
**Figura 4.7.** Curvas de enriquecimiento para un cribado virtual utilizando un modelo farmacofórico ideal (verde) y una selección de compuestos al azar (azul).

### 4.3. Base de datos.

#### 4.3.1. Base de datos de activos.

Se compila una base de datos de inhibidores del corresponsor de entrada CXCR4 descritos en la literatura y compuestos sintetizados en el grupo. En total se dispone de 151 compuestos de cuatro familias diferentes y con actividad anti-VIH inferior a 100  $\mu\text{M}$  (Tabla 4.2 y Anexo 6).

**Tabla 4.2.** Familias de compuestos de la base de datos de activos, estructura del compuesto más activo de cada familia y su actividad como inhibidor de CXCR4 expresada como  $EC_{50}$  en nM.

Familia	Número de compuestos	Compuesto más activo	$EC_{50}$ / nM
azamacrociclo	90	 105	0.8
KRH	13	 106	6.7
Dpa	10	 107	100
poliamina	38	 32{8,8}	19
total	151		

- Familia de compuestos del tipo azamacrociclo: 90 compuestos de tipo mono y bisazamacrociclo.<sup>[21-30]</sup>
- Familia de compuestos del tipo KRH: 13 análogos de KRH1636 desarrollados por Kureha Chemical Industry.<sup>[29,31-34]</sup>
- Familia de compuestos del tipo Dpa: 10 compuestos del tipo complejo de dipicolilamina y zinc(II).<sup>[35]</sup>
- Familia de compuestos del tipo poliamina: 38 compuestos sintetizados en el presente trabajo y descritos en el Capítulo 1.<sup>[9]</sup>

Todos estos compuestos se protonan a pH fisiológico, se les asignan las cargas parciales de Gasteiger y se optimiza su geometría utilizando el *force field* MMFF94.

### 4.3.2. Base de datos de inactivos.

#### 4.3.2.1. Inactivos seleccionados de la base de datos *Maybridge Screening Collection*.

Generalmente no se dispone de información sobre compuestos inactivos frente a una determinada diana terapéutica, y por lo tanto, el uso de bases de datos de presuntos inactivos es una práctica generalizada.<sup>[36]</sup>

Se han seleccionado 1462 compuestos presuntamente inactivos como inhibidores del VIH-1 de la base de datos *Maybridge Screening Collection*<sup>[37]</sup> que consta de 58,855 moléculas de tipo fármaco o *drug like* que generalmente cumplen la regla de Lipinski de los 5.<sup>[38]</sup>

La selección se basa en determinadas propiedades 1D de los compuestos, calculadas con MOE 2006.08:

- peso molecular (MW)
- número de átomos dadores de puente de hidrógeno (a\_acc)
- número de átomos aceptores de puente de hidrógeno (a\_don)
- número de enlaces simples rotables (b\_1rotN)
- número de átomos hidrofóbicos (a\_hyd)

Se incluyen aquellos inactivos que presentan valores de estas propiedades similares a los que presentan los compuestos que forman la base de datos de activos (Tabla 4.3 y Figura 4.8), de manera que se obtiene una base de datos focalizada para evitar enriquecimientos artificiales al cribar la base de datos de activos e inactivos con los farmacóforos modelados.<sup>[36]</sup>

Todos estos compuestos se protonan a pH fisiológico, se les asigna las cargas parciales de Gasteiger y se optimiza su geometría utilizando el *force field* MMFF94.

**Tabla 4.3.** Promedio y desviación estándar (entre paréntesis) de las propiedades 1D de los compuestos de la base de datos de activos y de inactivos.

Descriptor	Activos	Inactivos
MW	485.2 (104.9)	381.4 (64.9)
b_1rotN	6.9 (3.9)	5.1 (2.0)
a_acc	3.5 (1.5)	4.0 (1.1)
a_don	1.5 (1.5)	1.2 (1.1)
a_hyd	26.3 (5.0)	16.6 (2.6)

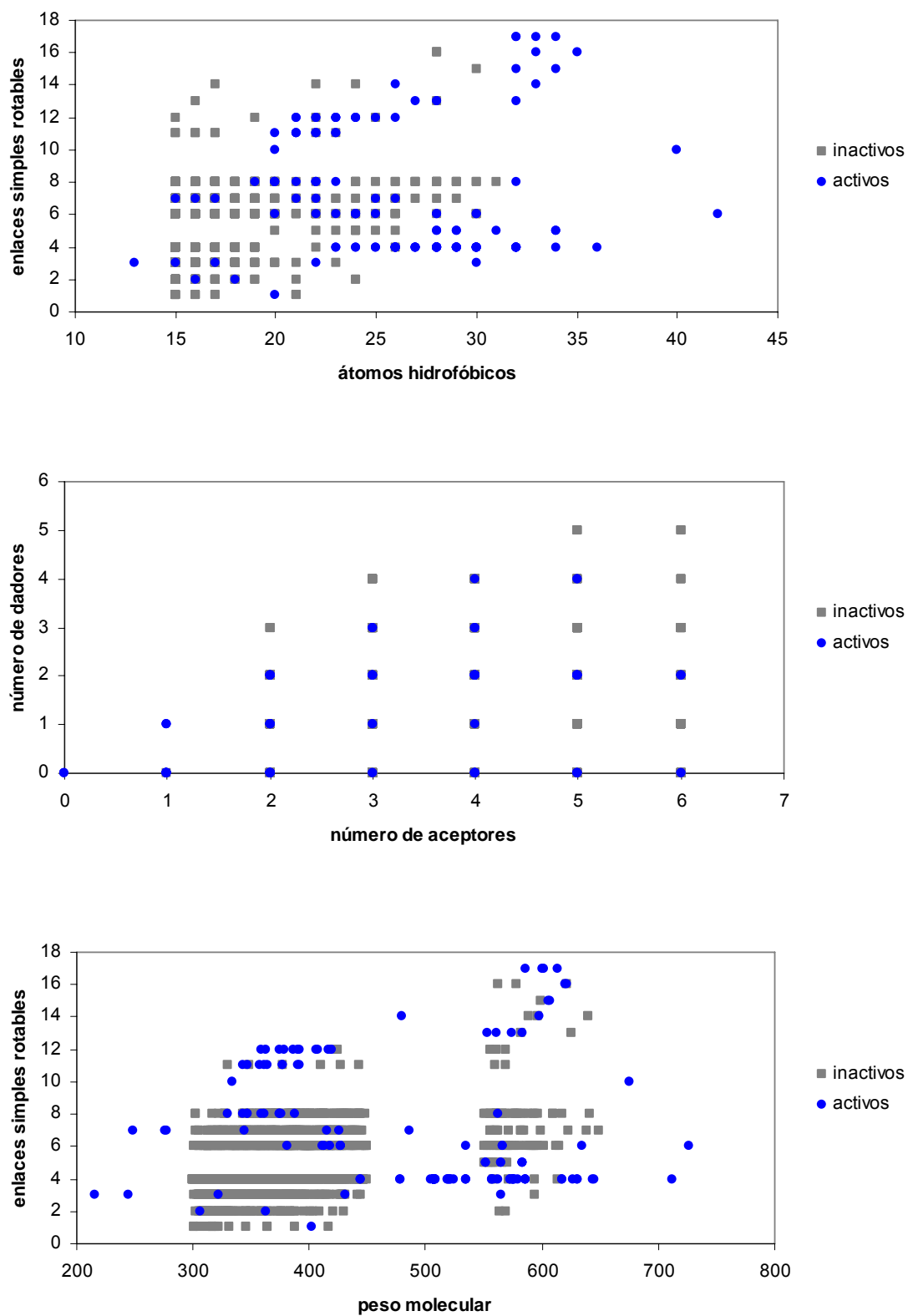


Figura 4.8. Propiedades de los compuestos activos e inactivos de la base de datos compilada.



#### 4.3.2.2. Inactivos reales.

Se seleccionan una serie de compuestos de la familia de las poliaminas, sintetizados en el Capítulo 1, que han demostrado ser inactivos como inhibidores del VIH-1 a concentraciones inferiores a 25  $\mu\text{g/mL}$  (Anexo 7). En este caso concreto se utiliza esta concentración como *cutoff* ya que es la máxima concentración utilizada en el grupo del Dr. Esté del IrsiCaixa al evaluar la actividad de los compuestos sintetizados en el presente trabajo. Estos compuestos se utilizan para evaluar el rendimiento de los modelos farmacofóricos frente a inactivos reales análogos a los activos de la familia de las poliaminas, quimioteca diseñada en el Capítulo 1. De esta manera, además de evaluar los farmacóforos mediante los parámetros habituales, es decir, los parámetros descritos anteriormente, se evalúa la capacidad de estos modelos de seleccionar compuestos activos frente a inactivos de una serie de análogos.

Todos estos compuestos se protonan a pH fisiológico, se les asignan las cargas parciales de Gasteiger y se optimiza su geometría utilizando el *force field* MMFF94.

#### **4.4. Alineamiento.**

La hipótesis o *query* farmacofórica se construye a partir de un conjunto de ligandos alineados entre sí. El objetivo de este alineamiento es que las características farmacofóricas similares estén próximas entre sí en el espacio 3D. Este proceso se realiza mediante el módulo *FlexAlign* de MOE<sup>[39]</sup> y se evalúa mediante los parámetros *F* (similitud), *U* (energía) y *S* (score) definidos en el Capítulo 3 (Ecuaciones 3.25-3.33).<sup>[40]</sup> Finalmente se selecciona el alineamiento que presenta el menor valor de *S*.

Se utilizan los siguientes alineamientos para modelar los farmacóforos:

1. AMD3100 (**24**), **106** (compuesto más activo de la familia KRH), y **32{8,8}** (compuesto más activo de la familia de poliaminas).
2. **105** (azamacrociclo más activo), **106**, y **32{8,8}**.
3. AMD3100 (**24**), **106**, **107** (compuesto tipo Dpa más activo), y **32{8,8}**.
4. **105**, **106**, **107**, y **32{8,8}** (Figura 4.9).



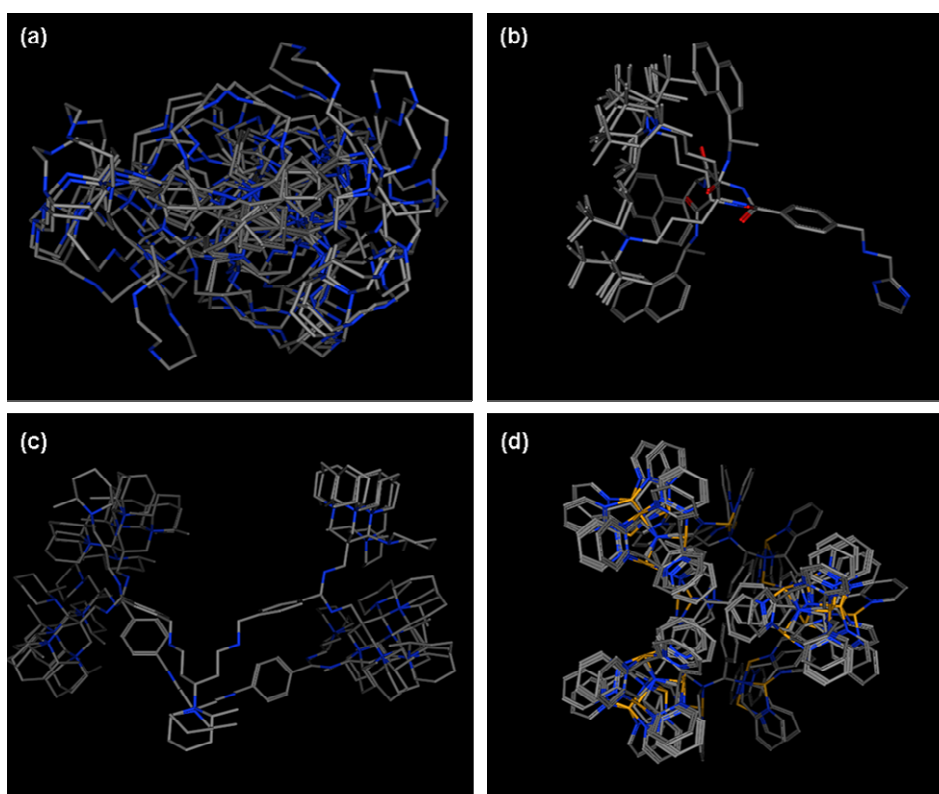
**Figura 4.9.** Alineamiento de menor *S* de los compuestos más activos de cada familia presente en la base de datos de activos: azamacrociclo **105** (lila), poliamina **32{8,8}** (azul), KRH **106** (verde) y Dpa **107** (gris).

#### **4.5. Cálculo de conformaciones.**

Para realizar un cribado virtual, ya sea un análisis retrospectivo como un análisis prospectivo, con un farmacóforo es necesario disponer de una base de datos de conformaciones para los compuestos en estudio.

Las conformaciones moleculares se pueden definir como estructuras que se pueden convertir de unas a otras simplemente por una rotación alrededor de un enlace sencillo.

Se calculan 50 conformaciones de menor energía para cada compuesto de la base de datos de activos e inactivos utilizando una búsqueda conformacional estocástica con el *force field* MMFF94, que cubre el espacio conformacional por muestreo al azar de los ángulos diedros en búsqueda de mínimos locales de la superficie de energía potencial.



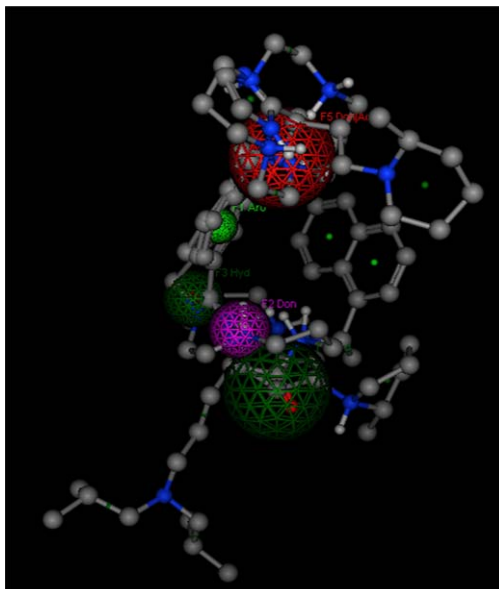
**Figura 4.10.** Conformaciones generadas para un compuesto representativo de cada familia de la base de datos de activos: (a) azamacrociclo, (b) KRH, (c) poliamina y (d) Dpa.

## **4.6. Modelos farmacofóricos: análisis retrospectivo.**

### **4.6.1. Modelo 1.**

Se construye la hipótesis farmacofórica a partir del alineamiento 1, es decir, la poliamina más activa **32**{8,8}, el compuesto tipo KRH más activo **106** y AMD3100 (azamacrociclo **24**), y se generan las *features* aplicando el esquema PCH con la herramienta *Pharmacophore Consensus* de MOE que sugiere características farmacofóricas comunes en un radio de tolerancia y un porcentaje determinado de los compuestos alineados.

El modelo obtenido consta de una *feature* aromática, una de dador de puente de hidrógeno, dos hidrofóbicas y una de aceptor o dador de puente de hidrógeno (Figura 4.11).



**Figura 4.11.** Representación del modelo farmacofórico 1 donde las features se representan con esferas coloreadas sobre los compuestos del alineamiento 1: Aro (verde claro), Don (lila), Hyd (verde oscuro) y Don|Acc (rojo).

El análisis retrospectivo de la base de datos de activos e inactivos permite estudiar el rendimiento del modelo farmacofórico, cuyo resultado se muestra en la Tabla 4.4. Este modelo clasifica incorrectamente 35 inactivos y 29 activos de la base de datos, presenta un rendimiento de activos del 78% y un *ratio* de activos del 81%.

**Tabla 4.4.** Análisis retrospectivo del modelo farmacofórico 1.

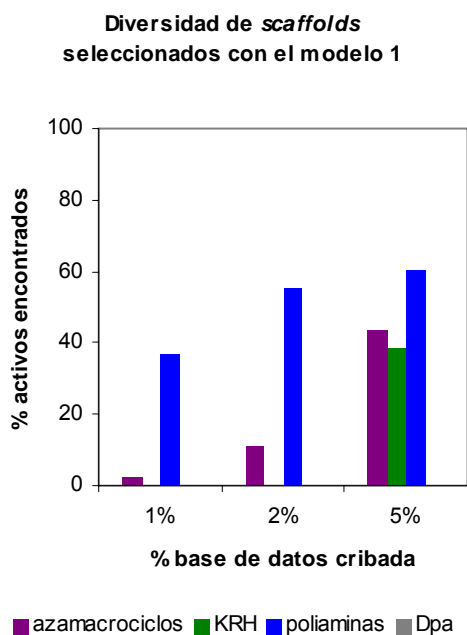
	$H_a$	$H_t$	falsos +	falsos -	<i>ef</i>	% Y	% A	<i>GH</i>	$GH_r$
<b>modelo 1</b>	122	157	35	29	8.30	77.71	80.79	0.77	0.09

Se analiza el factor de enriquecimiento obtenido a distintos porcentajes de base de datos cribada, el resultado se muestra en la Tabla 4.5. Al cribar un 1% de la base de datos se obtiene el factor de enriquecimiento máximo, y decrece al aumentar el porcentaje de base de datos cribada, al 5% su valor es de 8.95.

**Tabla 4.5.** Factor de enriquecimiento obtenido distintos porcentajes de base de datos cribada utilizando el modelo farmacofórico 1.

% base de de datos cribada	<i>ef</i>
1	10.68
2	10.35
5	8.95

También se lleva a cabo un análisis de diversidad de los *scaffolds* seleccionados por el modelo a distintos porcentajes de base de datos cribada (Figura 4.12). Se detecta que al 1% y 2% de base de datos cribada, el modelo selecciona preferentemente la familia de las poliaminas frente al resto de activos, seguido de la familia de los azamacrociclos, mientras que la familia de los compuestos tipo KRH no se selecciona hasta el 5% de base de datos cribada. No obstante, a este porcentaje, el modelo no exhibe ninguna preferencia clara sobre una familia determinada, aunque no selecciona ningún compuesto del tipo Dpa.

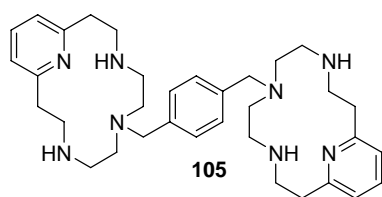


**Figura 4.12.** Análisis de diversidad de scaffolds al cribar la base de datos con el modelo 1.

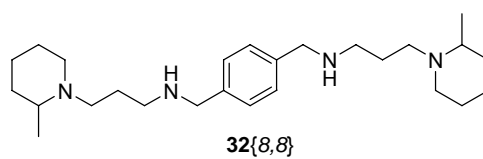
El análisis retrospectivo con el modelo farmacofórico 1 de los activos de la base de datos muestra que el modelo no ordena correctamente los compuestos, los más activos quedan posicionados hacia el centro y final del *hit list*, en especial los compuestos de la familia KRH que quedan todos posicionados al final cuando realmente deberían estar en las primeras posiciones. La familia de los azamacrociclos tampoco queda bien ordenada según su actividad. En cambio, la familia de las poliaminas queda correctamente posicionada en función de la actividad. Este modelo no clasifica como activo ninguna molécula de la familia de los azamacrociclos cuya estructura presenta sólo una unidad de ciclamo, dos macrociclos separados por un *linker* no aromático, cuyo macrociclo está complejado con metales, o compuestos de la familia Dpa (Figura 4.13).

Por otra parte, se realiza un análisis retrospectivo utilizando la base de datos de inactivos reales, es decir, compuestos de la familia de las poliaminas que no han resultado inactivos a concentraciones inferiores a 25  $\mu\text{g/mL}$ , solamente se clasifican correctamente como inactivos 3 de los 16 compuestos que forman esta base de datos.

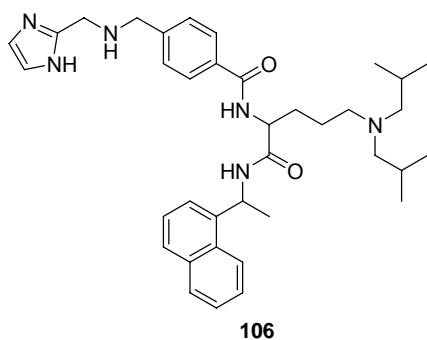
azamacrociclo



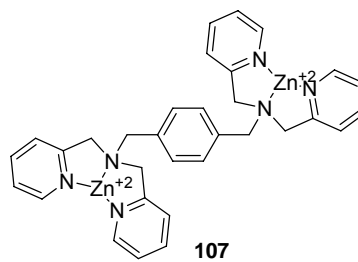
poliamina



KRH



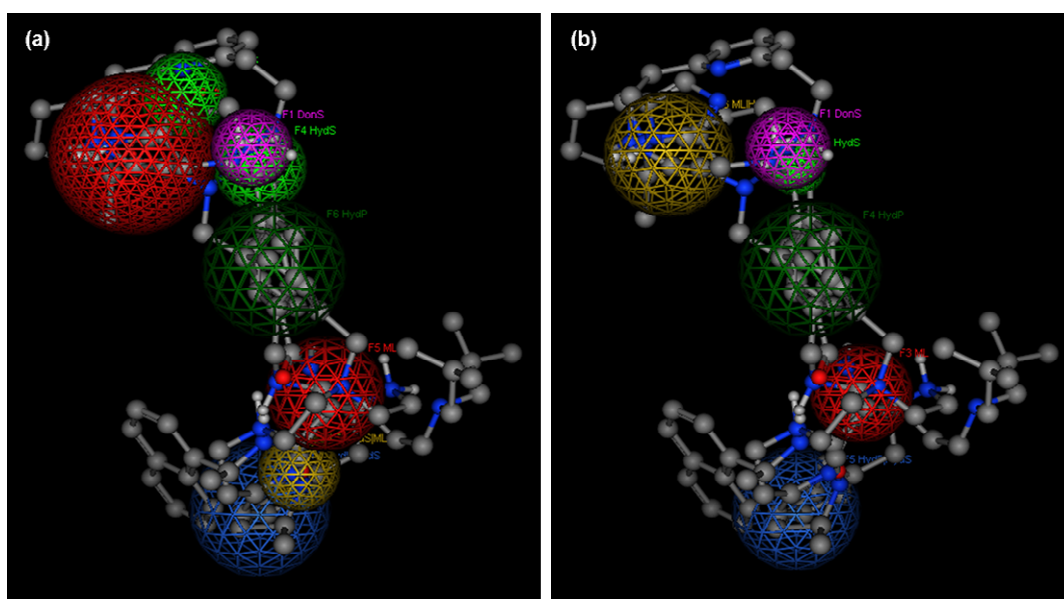
Dpa



**Figura 4.13.** Estructura de los compuestos más activos de cada familia.

### 4.6.2. Modelos 2 y 3.

Estos dos modelos farmacofóricos se generan a partir del alineamiento 2, es decir, **105** (azamacrociclo más activo), **106** (compuesto más activo de la familia KRH), y **32{8,8}** (poliamina más activa), se utiliza el esquema PPCH\_All para modelar las *features* y para la anotación de los ligandos con la herramienta *Pharmacophore Consensus*. La selección de unas u otras *features* sugeridas mediante este procedimiento permite obtener diferentes modelos entre los que destacan los presentados en la Figura 4.14, ambos más complejos que el farmacóforo 1.



**Figura 4.14.** (a) Modelo farmacofórico 2. (b) Modelo farmacofórico 3. En ambos las features se representan mediante esferas coloreadas sobre los compuestos del alineamiento 2: DonS (lila), HydS (verde claro), HydP (verde oscuro), HydP|HydS (azul). ML (rojo), y ML|HydS (amarillo).

Se realiza un análisis retrospectivo con ambos modelos de la base de datos de inactivos e activos. El resultado de los parámetros del análisis se resume en la Tabla 4.6. Se observa que el modelo 2 selecciona menos falsos positivos que el modelo 3, mientras que el número de falsos negativos es inferior en el modelo 3. El factor de enriquecimiento obtenido es superior en el modelo 2, así como el rendimiento de activos. Sin embargo, el ratio de activos obtenido con el modelo 3 es un 5% mayor que el obtenido con el modelo 2. La bondad del *hit list* obtenido con el farmacóforo 2 es 0.85, mayor que el obtenido con el modelo 3, cuyo valor es 0.78.



**Tabla 4.7.** Análisis retrospectivo de los modelos farmacofóricos 2 y 3.

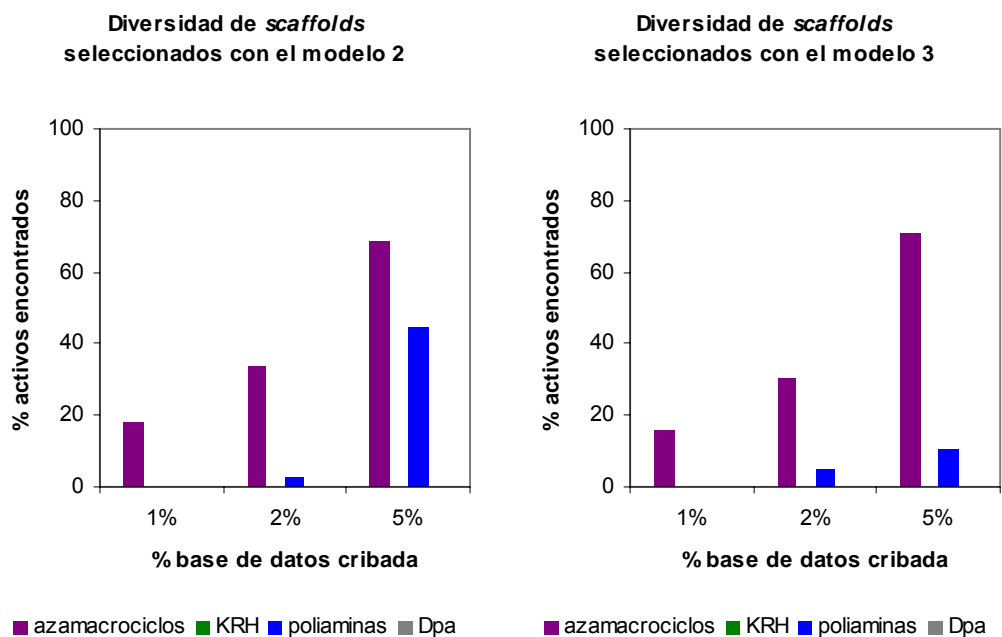
	$H_a$	$H_t$	falsos +	falsos -	$ef$	% Y	% A	GH	$GH_r$
<b>modelo 2</b>	121	137	16	30	9.43	88.32	80.13	0.85	0.08
<b>modelo 3</b>	128	163	35	23	8.39	78.53	84.77	0.78	0.09

El factor de enriquecimiento obtenido a distintos porcentajes iniciales de la base de datos cribada se detalla en la Tabla 4.7, donde se observa que el modelo 2 rinde mejores resultados ya que el factor de enriquecimiento es máximo al 1% de base de datos cribada y es superior en todos los casos al valor obtenido con el modelo 3.

**Tabla 4.8.** Factor de enriquecimiento obtenido a distintos porcentajes de base de datos cribada utilizando los modelos farmacofóricos 2 y 3.

% base de de datos cribada	modelo 2 $ef$	modelo 3 $ef$
1	10.68	9.35
2	10.35	9.68
5	10.55	9.08

El análisis de diversidad de las familias seleccionadas por los modelos farmacofóricos 2 y 3 se detalla en la Figura 4.15. En los porcentajes iniciales de base de datos cribada, el modelo farmacofórico 2 selecciona preferentemente los compuestos de la familia de los azamacrociclos. La siguiente familia en ser seleccionada es la de las poliaminas. Ni los compuestos tipo KRH ni Dpa se seleccionan en estos porcentajes iniciales de base de datos cribada con el modelo 2. El modelo 3 es aún más selectivo para la familia de los azamacrociclos, incluso al cribar hasta el 5% de la base de datos con los activos sin cargar solamente selecciona esta familia de compuestos. En general, estos modelos son selectivos para los azamacrociclos.



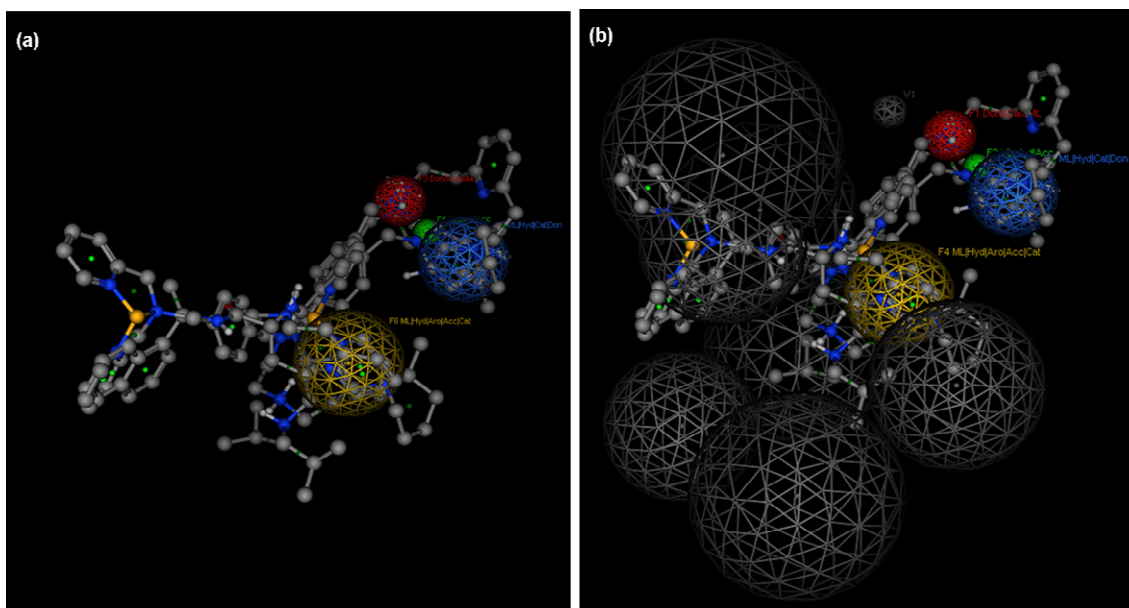
**Figura 4.15.** Análisis de diversidad de scaffolds al cribar la base de datos con los modelos 2 y 3.

En el análisis retrospectivo completo de la base de datos con el modelo 2, el *hit list* queda bien ordenado según la actividad de los ligandos, es decir, los más activos quedan posicionados al inicio de la lista mientras que los menos activos quedan posicionados al final, excepto la familia KRH que es seleccionada en último lugar. Este farmacóforo clasifica como activos los azamacrociclos complejados con metales. Sin embargo, no selecciona ni los azamacrociclos cuya estructura presenta solamente una unidad de ciclamo, ni cuyos azamacrociclos están separados por un *linker* no aromático. Tampoco considera activos los compuestos del tipo Dpa, ni las poliaminas con aldehído. El análisis retrospectivo completo de la base de datos con el modelo farmacofórico 3 rinde un *hit list* ordenado según la actividad de dichos compuestos, excepto la familia KRH que queda posicionada al final. Igual que en el caso anterior, este modelo selecciona como activos los azamacrociclos complejados con metales pero no aquellos formados solamente por una unidad de ciclamo ni cuyos azamacrociclos están separados por un *linker* no aromático. Tampoco selecciona compuestos tipo Dpa ni poliaminas con aldehído.

Si se utiliza el modelo farmacofórico 2 para cribar la base de datos de inactivos reales se obtiene una selección correcta de 8 de los 16 compuestos que la forman, mientras que el modelo 3 clasifica correctamente 7 inactivos reales.

### 4.6.3. Modelos 4 y 5.

Estos dos farmacóforos se modelan a partir del alineamiento 4, **105** (azamacrociclo más activo), **106** (compuesto más activo de la familia KRH), **107** (compuesto tipo Dpa más activo), y **32{8,8}** (poliamina más activa), con el esquema PCH para la anotación de los ligandos y selección de las *features*, y con la herramienta *Pharmacophore Consensus*. La adición de *features* de volúmenes de exclusión sobre el modelo 4 permite obtener el farmacóforo 5 (Figura 4.16).

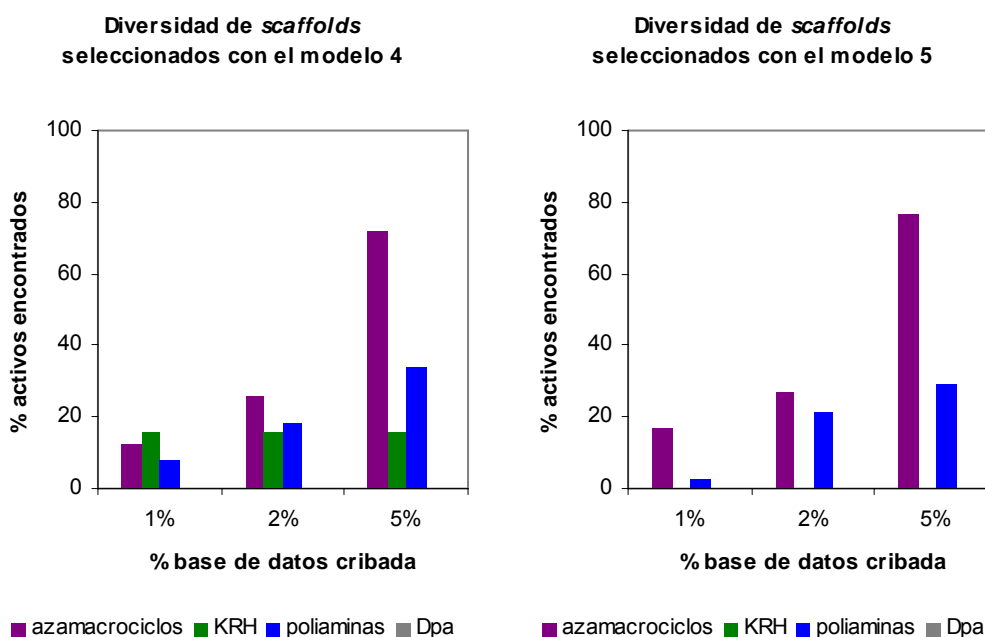


**Figura 4.16.** (a) Modelo farmacofórico 4. (b) Modelo farmacofórico 5. En ambos las features se representan mediante esferas coloreadas sobre los compuestos del alineamiento 4: Don&Cat&ML (rojo), ML|Hyd|Acc (verde), ML|Hyd|Cat|Don (azul), ML|Hyd|S|Aro|Acc|Cat (amarillo) y volumen de exclusión (gris).

Se realiza un análisis retrospectivo de ambos modelos farmacofóricos con la base de datos compilada para este propósito (Tabla 4.9). El resultado obtenido con la base de datos de activos cargados solamente rinde 1 falso positivo y 12 falsos negativos. Por consiguiente, el rendimiento de activos de estos dos modelos sobre la base de datos cribada es muy elevado, 99%, así como el ratio de activos, 92%. Además, el valor del factor de enriquecimiento de ambos modelos es muy elevado, 10.61 (el máximo que se puede obtener para la base de datos cribada dada su relación de compuestos activos e inactivos es 10.68). En resumen, se han obtenido unos modelos farmacofóricos muy selectivos.

**Tabla 4.9.** Análisis retrospectivo de los modelos farmacofóricos 4 y 5.

	$H_a$	$H_t$	falsos +	falsos -	ef	% Y	% A	GH	GH <sub>r</sub>
<b>modelo 4</b>	139	140	1	12	10.61	99.29	92.05	0.97	0.08
<b>modelo 5</b>	139	140	1	12	10.61	99.29	92.05	0.97	0.08



**Figura 4.17.** Análisis de diversidad de scaffolds al cribar la base de datos con los modelos 4 y 5.

La diversidad de familias seleccionadas por ambos modelos farmacofóricos se analiza en la Figura 4.17. Hasta el 5% de base de datos cribada, ninguno de los dos modelos selecciona la familia menos activa Dpa, y el modelo 5 tampoco selecciona compuestos del tipo KRH. Sin embargo, desde el inicio, este modelo selecciona las otras dos familias más activas de compuestos, preferentemente azamacrociclos pero también activos de la familia de las poliaminas. Por otra parte, al 1% de base de datos cribada, el modelo 4 ya selecciona tres familias de compuestos, azamacrociclos, poliaminas y del tipo KRH, pero muestra una clara preferencia por los azamacrociclos al 5% de base de datos cribada.

En el análisis retrospectivo del total de la base de datos utilizando el modelo farmacofórico 4, los compuestos no quedan ordenados según su actividad biológica, los más activos quedan posicionados en la zona central del *hit list*. Cabe destacar que este modelo

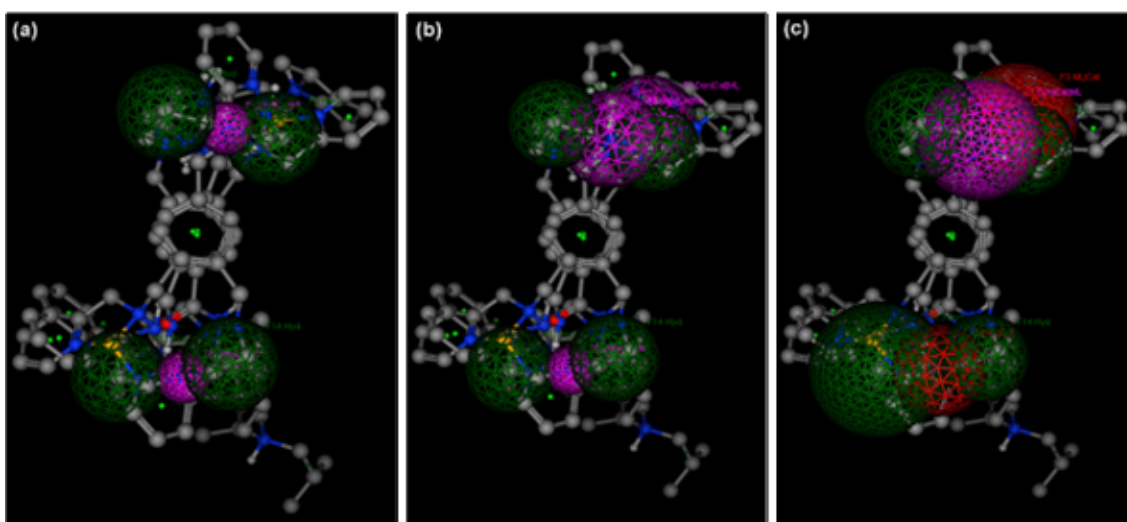
---

posiciona correctamente dos compuestos del tipo KRH, que quedan colocados en las dos primeras posiciones del *hit list*, también clasifica como activos el resto de compuestos de esta familia. Además, este farmacóforo y el farmacóforo 5 son capaces de seleccionar azamacrociclos constituidos por una sola unidad de ciclamo, cuyos macrociclos se hallan complejados con metales, y aquellos separados por un *linker* no aromático. Sin embargo, ninguno de los modelos clasifica como activos compuestos del tipo Dpa. El análisis retrospectivo global realizado con el modelo farmacofórico 5 tampoco ordena las moléculas según su actividad anti-VIH. A diferencia del modelo anterior, los compuestos del tipo KRH quedan posicionados al final del *hit list*, pero cabe destacar que los clasifica como activos.

La diferencia más notable entre estos dos modelos farmacofóricos es el resultado obtenido al cribar la base de datos de inactivos reales. El farmacóforo 4 clasifica correctamente como inactivos 6 compuestos de esta base de datos, mientras que los volúmenes de exclusión presentes en el modelo 5 permiten aumentar el número de inactivos correctamente clasificados a 9.

#### 4.6.4. Modelos 6, 7 y 8.

Estos tres farmacóforos se desarrollan a partir del alineamiento 3, AMD3100 (azamacrociclo **24**), **106** (compuesto más activo de la familia KRH), **107** (compuesto tipo Dpa más activo), y **32{8,8}** (poliamina más activa), y se aplica el esquema PCH (Figura 4.18). Las *features* se definen manualmente a partir de la interacción de los ligandos del tipo biciclamo con el correceptor de entrada CXCR4: interacción de los grupos carboxilatos de los ácidos aspárticos en las posiciones 171 y 262 del correceptor con los grupos NH de los anillos de ciclamo, interacción por puente de hidrógeno de los grupos NH de los ciclamos con el grupo carboxilato del ácido glutámico en posición 288, además de la coordinación de los metales complejados con ciclamos del carboxilato del ácido aspártico en posición 262 y las zonas hidrofóbicas presentes de los ligandos.<sup>[41-46]</sup>



**Figura 4.18.** (a) Modelo farmacofórico 6. (b) Modelo farmacofórico 7. (c) Modelo farmacofórico 8. Las *features* se representan mediante esferas coloreadas sobre los compuestos del alineamiento 3: Don|Cat|ML (lila), Hyd (verde) y Cat|ML (rojo).

En la Figura 4.18 se observa que el modelo farmacofórico 6 presenta una disposición simétrica de sus *features*. A partir de él, se obtiene el modelo 7, donde se aumenta el radio de la *feature* superior correspondiente a un dador de puente de hidrógeno, catión o ligando de metal (en lila) y además se añade otra *feature* con esta misma definición. El modelo 8 es el que presenta los mayores radios en las *features* de los tres farmacóforos, sobre todo los correspondientes a las zonas hidrofóbicas (en verde). Además, respecto al modelo 7 se han sustituido dos de las *features* de dador, catión o ligando de metal (en lila) por catión o ligando de metal (en rojo).

El resultado del análisis retrospectivo de la base de datos con los activos cargados y sin cargar a pH fisiológico con los modelos farmacofóricos 6-8 se resume en la Tabla 4.10. El modelo que presenta un mayor rendimiento de activos es el farmacóforo 7, mientras que el ratio de activos más elevado corresponde al modelo 8. El menor número de falsos negativos se obtiene con el farmacóforo 8, pero también rinde el mayor número de falsos positivos. Sin embargo, el mayor bondad de los *hit lists* se obtiene con el modelo 6. En cuanto al factor de enriquecimiento, el modelo 7 rinde el valor más elevado.

**Tabla 4.10.** Análisis retrospectivo de los modelos farmacofóricos 6-8.

	$H_a$	$H_t$	falsos +	falsos -	ef	% Y	% A	GH	GH <sub>r</sub>
<b>modelo 6</b>	133	186	53	18	7.64	71.51	88.08	0.73	0.09
<b>modelo 7</b>	123	168	45	28	7.82	73.21	81.46	0.73	0.09
<b>modelo 8</b>	141	245	104	10	6.15	57.55	93.38	0.62	0.09

El factor de enriquecimiento obtenido a distintos porcentajes iniciales de la base de datos cribada se detalla en la Tabla 4.11, donde se observa que hasta el 2% de base de datos cribada, en todos los casos se consigue el máximo factor de enriquecimiento posible, 10.68, y en el modelo 6 se mantiene hasta el 5%.

**Tabla 4.11.** Factor de enriquecimiento obtenido distintos porcentajes de base de datos cribada utilizando los modelos farmacofóricos 6, 7 y 8.

% base de de datos cribada	modelo 6 ef	modelo 7 ef	modelo 8 ef
1	10.68	10.68	10.68
2	10.68	10.68	10.68
5	10.68	9.75	9.75
10	8.43	7.83	8.03

El análisis de diversidad de las familias seleccionadas por los modelos 6-8 en los primeros porcentajes de base de datos cribada se representa en la Figura 4.19. Todos los modelos seleccionan todas las familias de activos en el 10% de base de datos cribada. Además en el análisis de la base de datos con cada uno de los farmacóforos modelados, en el 10% de base de datos cribada se incluye en el *hit list* la totalidad de los compuestos tipo KRH y

prácticamente todos los azamacrociclos. Sin embargo, en todos los casos se observa una clara preferencia por los activos de la familia de los azamacrociclos.

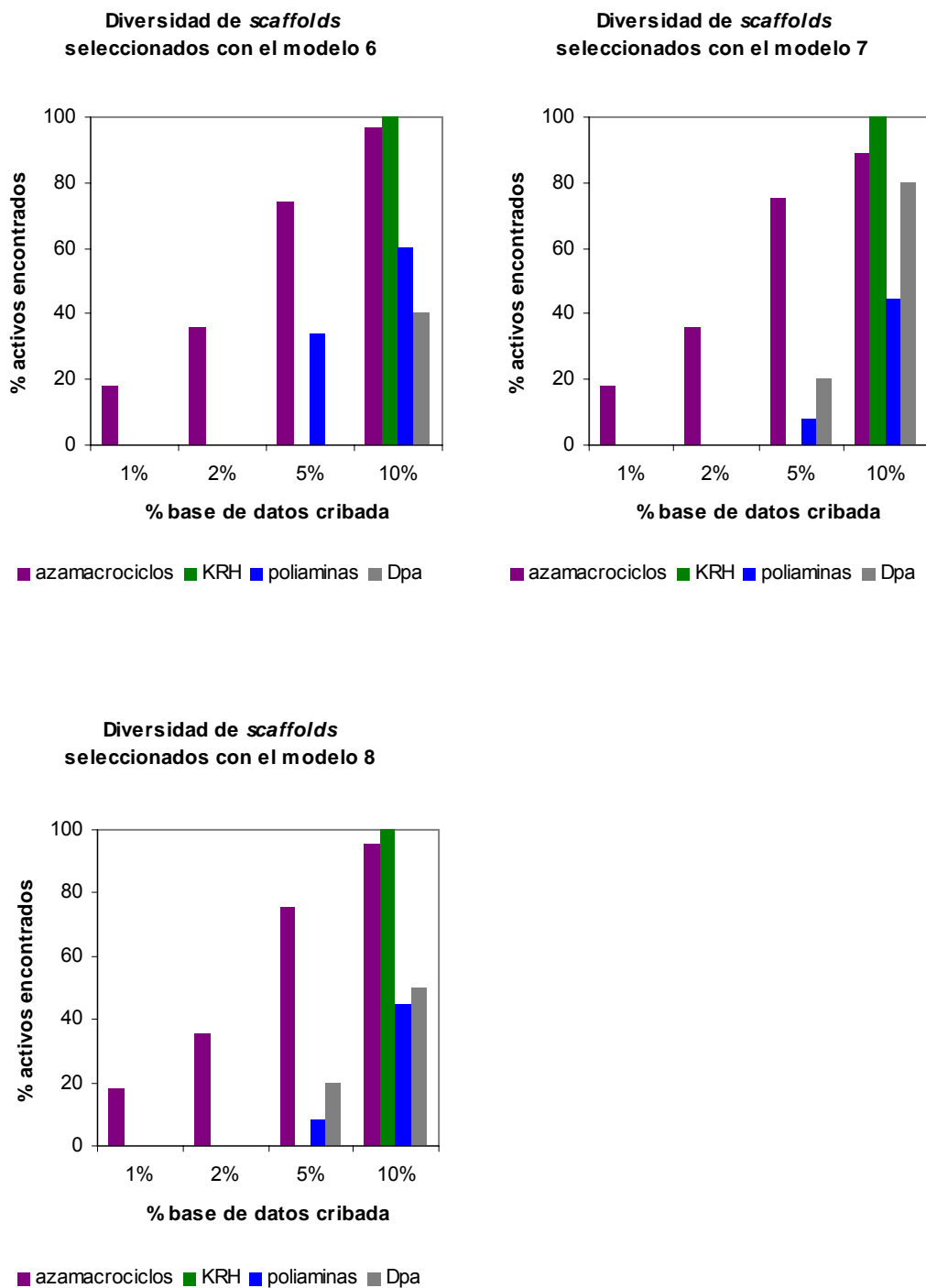


Figura 4.19. Análisis de diversidad de scaffolds al cribar la base de datos con los modelos 6, 7 y 8.



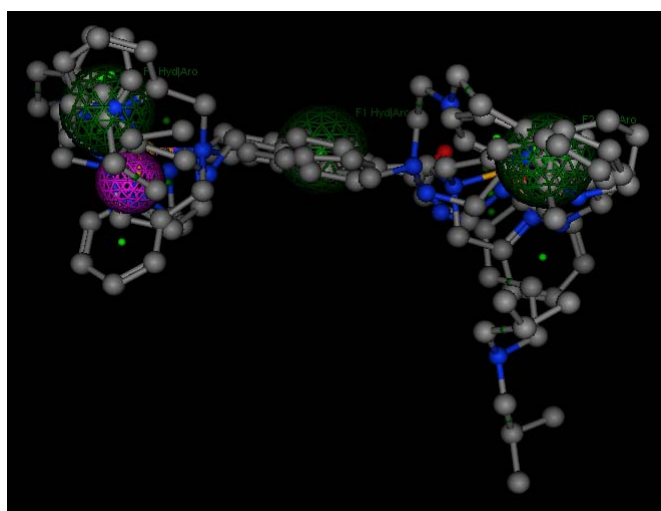
En el análisis retrospectivo del total de la base de datos con el modelo farmacofórico 6, los compuestos quedan ordenados según su actividad biológica, excepto los compuestos tipo KRH que quedan posicionados al final del *hit list* en vez de al inicio como les corresponde. En este análisis también se detectan como activos compuestos del tipo Dpa, azamacrociclos cuya estructura está formada por una sola unidad de ciclamo, cuyos macrociclos presentan un espaciador no aromático y aquellos complejados con metales. Sin embargo, clasifica incorrectamente los compuestos de la familia de las poliaminas con grupo aldehído. El mismo análisis con el modelo farmacofórico 7 presenta resultados análogos a los expuestos para el modelo 6. Sin embargo, el análisis retrospectivo con el modelo 8 rinde muchos falsos negativos de la familia de las poliaminas.

El análisis retrospectivo de la base de datos de inactivos reales con el modelo 6 proporciona 9 compuestos correctamente asignados como inactivos, mientras que el modelo 7 clasifica correctamente 10 inactivos reales y el modelo 8 clasifica como inactivos 8 de los 16 compuestos que forman esta base de datos.

En general, los modelos que presentan mejores resultados globales son los farmacóforos 6 y 7.

### 4.6.5. Modelo 9.

Se construye la hipótesis farmacofórica a partir de **105** (azamacrociclo más activo), **106** (compuesto más activo de la familia KRH), **107** (compuesto tipo Dpa más activo), y **32{8,8}** (poliamina más activa), con el esquema PCH y mediante la herramienta *Pharmacophore Elucidate* de MOE. Esta herramienta genera hipótesis farmacofóricas, por incremento sucesivo hasta un máximo de 5 *features*, a partir del análisis conformacional de los ligandos especificados. Se selecciona aquella hipótesis que presenta una mayor superposición con estos compuestos (Figura 4.20).



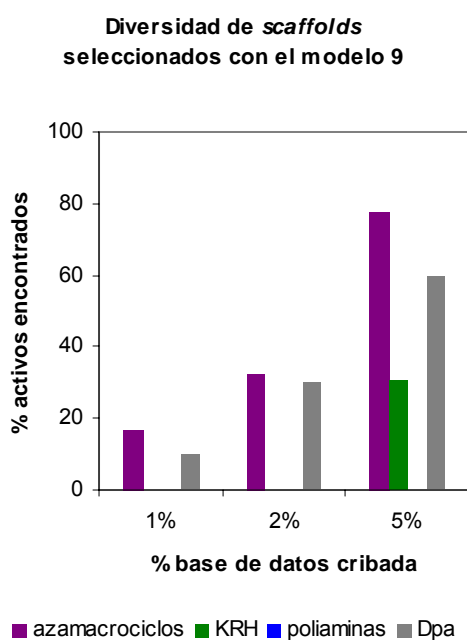
**Figura 4.20.** Representación del modelo farmacofórico 9 donde las features se representan con esferas coloreadas sobre los compuestos del alineamiento 4: Hyd|Aro (verde) y Cat (lila).

Los parámetros del análisis retrospectivo de la base de datos con este modelo farmacofórico se detallan en la tabla 4.12. Al cribar la base de datos se obtienen 142 compuestos en el *hit list*, todos ellos activos reales, de manera que el modelo no presenta ningún falso positivo. Además sólo clasifica incorrectamente como inactivos 9 ligandos, aunque 5 de ellos presentan valores de  $EC_{50}$  inferiores a 1  $\mu$ M. A partir de estos datos se calcula el rendimiento de activos que en este caso es del 100% y con un factor de enriquecimiento máximo, 10.68.

**Tabla 4.12.** Análisis retrospectivo del modelo farmacofórico 9.

	$H_a$	$H_t$	falsos +	falsos -	$ef$	% Y	% A	$GH$	$GH_r$
<b>modelo 9</b>	142	142	0	9	10.68	100.00	94.04	0.99	0.08

El análisis de diversidad de las familias seleccionadas por el modelo farmacofórico 9 en los primeros porcentajes de base de datos cribada se representa en la Figura 4.21. Se observa que al 1 y 2% de base de datos cribada se seleccionan azamacrociclos y compuestos tipo Dpa, y no es hasta el 5% que se detectan activos tipo KRH. Sin embargo, hasta el 5% de base de datos cribada no se selecciona ningún compuesto de la familia de las poliaminas.



**Figura 4.21.** Análisis de diversidad de *scaffolds* al cribar la base de datos con el modelo 9.

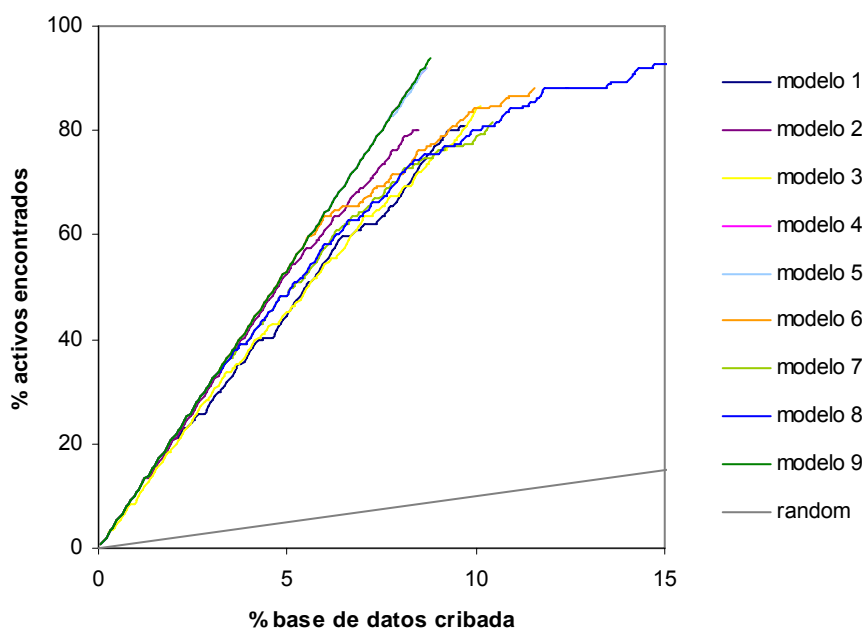
El *hit list* obtenido al realizar el análisis retrospectivo del total de la base de datos de activos e inactivos utilizando el modelo farmacofórico 9, presenta los ligandos correctamente ordenados según su actividad anti-VIH, excepto los activos tipo KRH que deberían posicionarse al inicio de la lista. Este modelo selecciona activos de la familia Dpa, así como azamacrociclos cuyos ciclamos se hallan complejados con metales o separados por un *linker* no aromático. Sin embargo clasifica incorrectamente los azamacrociclos que presentan una sola unidad de ciclamo en su estructura.

El cribado virtual de la base de datos de inactivos reales con el farmacóforo 9 clasifica correctamente como inactivos 9 de los 16 compuestos que forman esta base de datos.

En resumen, el modelo 9 es el farmacóforo que presenta un mejor rendimiento, Y, al realizar un análisis retrospectivo con la base de datos de activos e inactivos compilada.

#### 4.6.6. Comparación de los modelos farmacofóricos.

En primer lugar se representan las curvas de enriquecimiento obtenidas al realizar el cribado virtual de la base de datos de activos e inactivos con cada uno de los farmacóforos modelados, y se comparan con el resultado de una selección *random* (Figura 4.22). Se observa que todos los modelos presentan una selección de activos mejor que la selección al azar. Destaca el resultado obtenido al cribar la base de datos con el modelo 9.



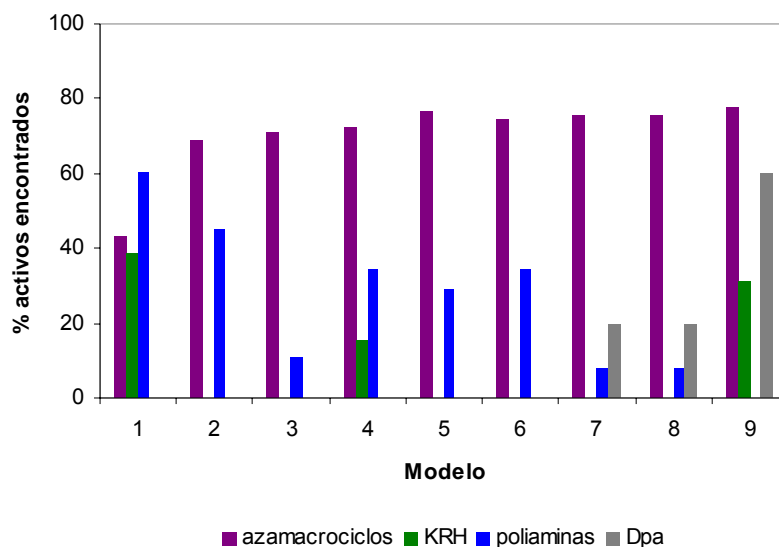
**Figura 4.22.** Curvas de enriquecimiento obtenidas al cribar la base de datos de activos e inactivos con los farmacóforos modelados.

Se comparan los resultados obtenidos en el análisis retrospectivo de los diferentes modelos (Tabla 4.13). Se observa que los modelos farmacofóricos que han presentado mejores valores en los parámetros del análisis retrospectivo de la base de datos de activos e inactivos compilada son los modelos 4, 5 y 9, aunque el modelo 5 es menos diverso que los otros dos modelos en la selección de las diferentes familias de ligandos.

**Tabla 4.13.** Comparación del resultado del análisis retrospectivo de los diferentes modelos farmacofóricos, los mejores valores obtenidos de cada parámetro se resaltan en negrita.

Modelo	ef	falsos +	falsos -	%Y	%A
1	8.30	35	29	77.71	80.79
2	<b>9.43</b>	<b>16</b>	30	<b>88.32</b>	80.13
3	8.39	35	23	78.53	84.77
4	<b>10.61</b>	<b>1</b>	<b>12</b>	<b>99.29</b>	<b>92.05</b>
5	<b>10.61</b>	<b>1</b>	<b>12</b>	<b>99.29</b>	<b>92.05</b>
6	7.64	53	18	71.51	88.08
7	7.82	45	28	73.21	81.46
8	6.15	104	<b>10</b>	57.55	<b>93.38</b>
9	<b>10.68</b>	<b>0</b>	<b>9</b>	<b>100.00</b>	<b>94.04</b>

Otro parámetro de comparación del rendimiento de los farmacóforos modelados es el porcentaje de compuestos de cada familia de activos que cada uno de los modelos es capaz de seleccionar en los primeros porcentajes de base de datos de activos e inactivos cribada. Un modelo versátil debe seleccionar el mayor número de compuestos de cada familia, cuanto más similares sean el porcentaje de compuestos obtenidos para cada familia, más general será el modelo. En cambio, un modelo que seleccione un gran número de compuestos de una sola familia, pero no del resto de familias, será un modelo selectivo para aquella familia de ligandos. La comparación de las familias seleccionadas de los diferentes modelos farmacofóricos desarrollados, al 5% de base de datos de activos e inactivos cribada, se resume en la Figura 4.23. Se observa que todos los modelos son altamente selectivos para los azamacrociclos, en casi todos los modelos se selecciona más de un 60% de moléculas de esta familia. En general, también muestran preferencia por las poliaminas, los modelos 1, 2, 4, 5 y 6 seleccionan más de un 25% de compuestos de esta familia. Por otra parte, el modelo 1 selecciona el mayor porcentaje de moléculas del tipo KRH, pero no detecta ningún activo del tipo Dpa, mientras que el modelo 9 selecciona un 60% de compuestos de esta familia pero no incluye en su *hit list* ninguna poliamina. El modelo 4, aunque demuestra preferencia por los compuestos del tipo azamacrociclo, también selecciona más de un 30% de compuestos de la familia de las poliaminas, y prácticamente un 15% de compuestos del tipo KRH. La conclusión de este análisis es que la combinación de los modelos farmacofóricos 1 y 9 o de los modelos 4 y 9 rinde la mayor diversidad de compuestos seleccionados al realizar un cribado virtual de la base de datos de activos e inactivos.



**Figura 4.23.** Análisis de diversidad de scaffolds seleccionados al 5% de base de datos cribada para los farmacóforos modelados.

Si se analiza el resultado obtenido en el análisis retrospectivo del total de la base de datos de activos e inactivos con los diferentes farmacóforos modelados, se observa que en general los compuestos del tipo KRH quedan posicionados al final de la lista de activos excepto con el modelo 4. Sin embargo, este modelo no ordena correctamente según la actividad antiviral los azamacrociclos y las poliaminas. No obstante, estas dos familias de compuestos están correctamente posicionados en los *hit lists* obtenidos con los modelos 2, 3, 6, 7, 8 y 9. Por otra parte, los compuestos del tipo Dpa solamente se identifican como activos al realizar el análisis retrospectivo de la base de datos con los modelos 6-9, aunque representa la familia de ligandos de menor actividad de las cuatro familias que componen la base de datos de activos. Los farmacóforos 1-3 clasifican como inactivos azamacrociclos cuyos macrociclos están separados por un *linker* no aromático y aquellos formados por un solo macrociclo. Además, este último subgrupo también se identifica como inactivo con el modelo 9. Otro subgrupo incorrectamente clasificado es el de las poliaminas con presencia de aldehído, ya que los farmacóforos 2, 3, 6, 7 y 8 lo etiquetan como inactivo.

Finalmente, se compara el rendimiento de los farmacóforos sobre la base de datos de inactivos reales de la familia de las poliaminas, ya que el análisis prospectivo se realizará sobre compuestos análogos a los compuestos de esta familia de activos. Aquellos que han presentado un mejor resultado en este análisis son los modelos 5, 6 y 9, que clasifican correctamente 9 inactivos reales, y el modelo farmacofórico 7 clasifica correctamente 10 de los 16 inactivos que componen esta base de datos.

Los resultados generales del análisis retrospectivo de los farmacóforos desarrollados muestran que ninguno de los modelos obtenidos permite obtener un resultado totalmente satisfactorio para todas las familias de compuestos, sin embargo, todos ellos presentan buenos

resultados generales. Cabe destacar que los modelos que presentan el mejor rendimiento son los farmacóforos 4 y 9. En el análisis prospectivo se utilizan todos los farmacóforos desarrollados y la decisión de los compuestos a sintetizar se toma por consenso de los resultados obtenidos con cada uno de estos modelos, los modelos QSAR y otros modelos basados en los ligandos (*shape matching*) y basados en el receptor modelado (*docking*) desarrollados en el grupo.<sup>[47]</sup>