

UNIVERSIDAD NACIONAL DE LA PLATA  
FACULTAD DE CIENCIAS EXACTAS  
DEPARTAMENTO DE CIENCIAS BIOLÓGICAS



Topología molecular aplicada al  
reconocimiento de sustratos de la  
Proteína de Resistencia del  
Cáncer de Mama (BCRP)

---

**Trabajo de Tesis Doctoral**

Farm. Melisa Edith Gantner

*Director:* Dr. Alan Talevi

2016





**UNIVERSIDAD NACIONAL DE LA PLATA**  
**FACULTAD DE CIENCIAS EXACTAS**  
**DEPARTAMENTO DE CIENCIAS BIOLÓGICAS**

---

Trabajo de Tesis Doctoral

**TOPOLOGÍA MOLECULAR APLICADA AL  
RECONOCIMIENTO DE SUSTRATOS DE LA  
PROTEÍNA DE RESISTENCIA DEL CÁNCER  
DE MAMA (BCRP)**

---

Farm. Melisa Edith Gantner

Director: Dr. Alan Talevi

2016



El presente trabajo de tesis ha sido desarrollado para acceder al grado académico de Doctor de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata. El mismo se realizó en la Cátedra de Química Medicinal, actual Laboratorio de Investigación y Desarrollo de Bioactivos (LIDeB), de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata bajo la dirección del Dr. Alan Talevi, y fue posible gracias a las becas otorgadas por la Comisión de Investigaciones Científicas de La Provincia de Buenos Aires (CIC-PBA) y por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).



A mis abuelos, Angela y Abraham

A mis padres, Marta y Eduardo



## *Agradecimientos*

A mi director Dr. Alan Talevi, por haber confiado en mí, haberme guiado y acompañado en este camino, por sus consejos, su ejemplo y su invaluable ayuda.

Al Dr. Luis Enrique Bruno-Blanch, por sus consejos y su gran ayuda en todo momento, por haber creído en mí y haberme abierto las puertas de su laboratorio.

A la Dra. Luciana Gavernet, por su amistad, charlas, consejos y ayuda siempre que lo necesité.

A la Dra. María Esperanza Ruiz, por su amistad, por estar siempre al pie del cañón para ayudarme, entablar colaboraciones y por ser quien me guió y ayudo en las determinaciones analíticas y estadísticas de este trabajo.

A la Dra. Roxana Peroni, por su guía y enorme disposición para trabajar en colaboración y ayudarme a realizar los ensayos experimentales de esta tesis.

A la Facultad de Ciencias Exactas, la UNLP, la CIC y el CONICET por brindarme los medios para llevar a cabo esta tesis.

A todos mis compañeros y amigos del LIDeB, por haber hecho de este un viaje maravilloso, tanto que he decidido quedarme a trabajar con todos ellos.

A mis compañeros de docencia, por tantas horas de trabajo y diversión tratado de transmitir algún conocimiento.

A mi familia; mis padres Marta y Eduardo, mi hermana Erica y mi novio José, son mi ancla en esta vida... sin ellos nada sería posible.

A mis amigos y hermanos de la vida, esos que están siempre y que le dan sentido a todo.



# Índice

---

<b>Objetivo</b> .....	1
<b>Capítulo I: Introducción</b> .....	<b>3</b>
1.1 Resistencia a múltiples fármacos.....	5
1.2 Transportadores de eflujo ABC .....	7
1.2.1 Un poco de historia .....	7
1.2.2 Características estructurales .....	9
1.2.3 Función .....	15
1.2.4 Mecanismo de acción .....	17
1.3 Transportadores MDR-ABC .....	20
Referencias Capítulo I.....	21
<b>Capítulo II: Epilepsia refractaria</b> .....	<b>27</b>
2.1 Definiciones.....	29
2.2 Hipótesis de la epilepsia refractaria .....	31
2.3 Hipótesis de los transportadores: la evidencia .....	36
2.3.1 Evidencia preclínica.....	37
2.3.2 Evidencia clínica.....	39
2.3.3 Transportadores MDR-ABC: un trabajo en equipo.....	40
2.4 Estrategias para superar la MDR mediada por transportadores ABC en la epilepsia refractaria.....	44
2.4.1 Aproximación aplicada en la tesis .....	48
Referencias Capítulo II .....	50
<b>Capítulo III: Modelos QSAR</b> .....	<b>63</b>
3.1 Descubrimiento de fármacos: breve reseña histórica .....	65
3.2 Métodos computacionales en el descubrimiento de fármacos .....	69
3.2.1 Cribado Virtual (VS) .....	71

3.3 Modelos de relación estructura-actividad cuantitativa .....	76
3.3.1 QSAR desde una perspectiva histórica .....	76
3.3.2 Construcción de un modelo QSAR.....	78
3.4 QSAR aplicado al cribado ADMET in silico.....	80
Referencias Capítulo III.....	82
<b>Capítulo IV: Materiales y métodos.....</b>	<b>91</b>
4.1 Conjunto de datos .....	93
4.2 Partición del conjunto de datos en los conjuntos de entrenamiento y de prueba .....	96
4.3 Cálculo de los descriptores moleculares.....	101
4.4 Método de modelado .....	102
4.4.1 Primera instancia de modelado: construcción de modelos lineales .....	104
4.4.1.1 Primera serie de modelos lineales .....	104
4.4.1.2 Segunda serie de modelos lineales .....	107
4.4.2 Segunda instancia de modelado: construcción de modelos no lineales ..	109
4.5 Evaluación del desempeño de los modelos.....	119
4.6 Validación in silico de los modelos.....	121
4.6.1 Validación interna .....	121
4.6.2 Validación externa.....	122
4.6.2.1 Biblioteca simulada .....	123
4.6.2.2 Biblioteca de señuelos DUD-E.....	123
4.7 Combinación de modelos: fusión de datos.....	124
4.7.1 Combinación de los modelos lineales .....	125
4.7.2 Combinación de los modelos no lineales .....	126
4.8 Selección de los mejores modelos y/o combinaciones .....	126
4.9 Validación experimental de la mejor combinación de modelos obtenida	127
4.9.1 Selección de los candidatos a evaluar experimentalmente.....	127
4.9.2 Evaluación experimental de los compuestos seleccionados .....	128

4.9.3 Análisis cuantitativo de las muestras por cromatografía líquida de alta resolución (HPLC).....	133
4.9.4 Análisis estadístico de los resultados experimentales.....	134
Referencias Capítulo IV.....	135
<b>Capítulo V: Resultados .....</b>	<b>147</b>
5.1 Partición del conjunto de datos en conjunto de entrenamiento y de prueba .....	149
5.2 Modelos lineales .....	150
5.2.1 Primera serie de modelos lineales.....	150
5.2.2 Segunda serie de modelos lineales.....	158
5.3 Modelos No Lineales .....	164
5.4 Validación experimental de la combinación seleccionada.....	172
5.4.1 Selección de los compuestos a evaluar .....	172
5.4.2 Condiciones establecidas para la cuantificación por HPLC.....	173
5.4.3 Validación del modelo experimental.....	174
5.4.4 Evaluación de las drogas según protocolo 1 .....	176
5.4.5 Evaluación de las drogas según protocolo 2 .....	179
Referencias Capítulo V.....	183
<b>Capítulo VI: Conclusiones .....</b>	<b>185</b>
<b>Anexo I.....</b>	<b>193</b>
A.1.1 Estructuras químicas de los compuestos del conjunto de datos .....	195
A.1.1.1 Estructuras químicas de los sustratos de la BCRP del conjunto de entrenamiento.....	195
A.1.1.2 Estructuras químicas de los no sustratos de la BCRP del conjunto de entrenamiento.....	203
A.1.1.3 Estructuras químicas de los sustratos de la BCRP del conjunto de prueba .....	210
A.1.1.4 Estructuras químicas de los no sustratos de la BCRP del conjunto de prueba.....	217

A.1.2 Estructuras químicas de la base de datos de 21 compuestos anticonvulsivos identificados por el LIDeB sobre la que se aplicó la mejor combinación de modelos obtenida .....	219
<b>Anexo II .....</b>	<b>221</b>
A.2.1 Ecuaciones de la segunda serie de modelos lineales.....	223
A.2.2 Diagramas de los 12 árboles de decisión obtenidos.....	225

## ***Objetivo general del trabajo de tesis:***

---

El objetivo general del plan de tesis es el desarrollo de modelos computacionales químico-matemáticos basados en descriptores moleculares topológicos capaces de discriminar entre sustratos y no sustratos de la Proteína de Resistencia del Cáncer de Mama (*Breast Cancer Resistance Protein*, BCRP/ABCG2), utilizando metodologías de modelado desde el ligando. Dichos modelos constituirán filtros secundarios *in silico* que podrán aplicarse en campañas de cribado virtual orientadas al descubrimiento de fármacos para el tratamiento de patologías con alta prevalencia de fenómenos de resistencia a múltiples fármacos asociados al transportador BCRP como la epilepsia y el cáncer, y para la predicción de potenciales reacciones medicamentosas debidas a la competencia de dos fármacos por dicho transportador.

## ***Objetivos específicos:***

---

- ∅ Compilación de un conjunto de datos compuesto por sustratos y no sustratos de la BCRP, de la que se obtendrán los conjuntos de entrenamiento y conjunto de prueba de los modelos a desarrollar.
- ∅ Obtención de los modelos capaces de discriminar entre sustratos y no sustratos de la BCRP, mediante aplicación de distintas metodologías tanto para la selección de las variables independientes (descriptores moleculares) como para la construcción y caracterización del modelo.
- ∅ Obtención de combinaciones de modelos individuales mediante la aplicación de esquemas de fusión de datos.
- ∅ Construcción de bibliotecas de mayor tamaño para una adecuada validación externa *in silico* de los modelos individuales y combinaciones generados.
- ∅ Cálculo de las curvas ROC (Característica Operativa del Receptor) para establecer la capacidad de clasificación de los modelos individuales y combinaciones generadas.
- ∅ Validación experimental del mejor modelo obtenido.



# Capítulo I

## Introducción

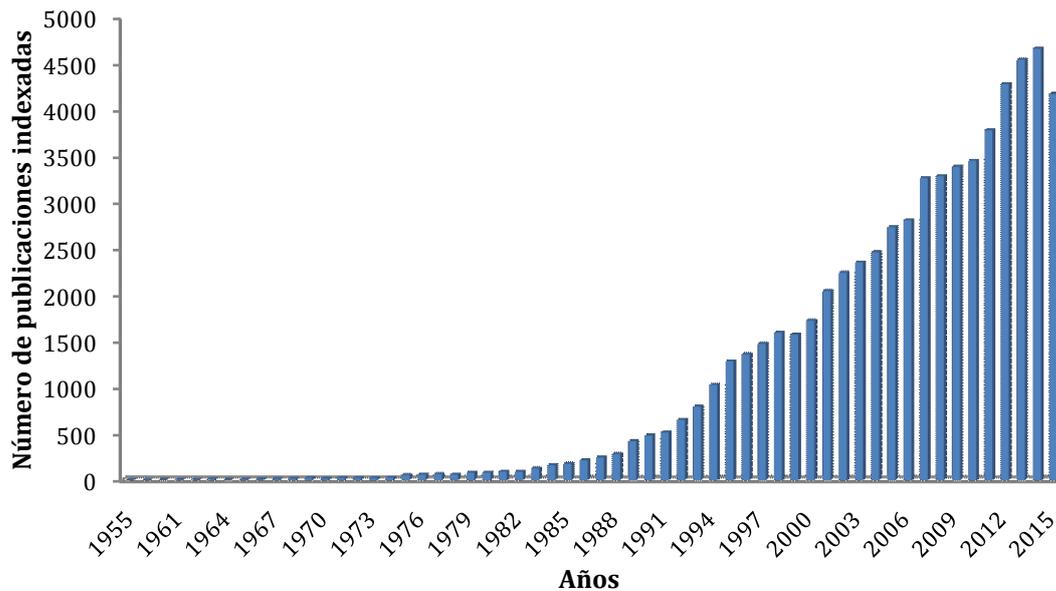
---



## 1.1 Resistencia a múltiples fármacos

La disminución de la eficacia de un fármaco en el tratamiento o cura de una determinada enfermedad o condición se conoce como resistencia a los fármacos. La resistencia a múltiples fármacos o MDR (por sus siglas en inglés, *multidrug resistance*) se puede definir, en términos generales, como la capacidad de una célula viva u organismo de ofrecer resistencia al tratamiento frente a un amplio espectro de fármacos no relacionados estructural o funcionalmente entre sí [1].

Las primeras nociones de MDR comenzaron a reportarse a mediados de la década de 1950, donde se analizaba la resistencia a múltiples fármacos desarrollada por *Mycobacterium tuberculosis* [2]. El concepto de MDR en tumores se remonta a finales de la década de 1960 y principios de 1970, cuando comenzó a reportarse un fenómeno de resistencia cruzada a múltiples fármacos de origen natural en líneas celulares seleccionadas a partir de uno de esos fármacos [3-5]. A comienzos de 1970, June Biedler identificó un amplio perfil de resistencia cruzada en células de pulmón de hámsters chinos seleccionados por presentar resistencia frente a actinomicina D [6], mientras que otro grupo de investigación reportaba paralelamente similares hallazgos frente a daunorubicina [7]. Nada hubiera permitido predecir en aquel entonces que poco más de medio siglo después, serían recuperadas de PubMed más de 65000 publicaciones cuando se introduce el simple término “*multidrug resistance*” (Figura 1.1). La MDR puede ocurrir en toda célula viva u organismo, por lo que el alcance de este fenómeno es enorme; el avance en el conocimiento de la MDR ha ido develando su complejidad, importancia e impacto en la clínica, generando un progresivo aumento del interés en la comunidad, especialmente durante los últimos 15 años.



**Figura 1.1.** Evolución del número de publicaciones indexadas hasta fines de 2015 en PubMed que se obtienen introduciendo “*multidrug resistance*” como palabra clave para la búsqueda.

Los mecanismos por los cuales una célula u organismo presenta MDR son múltiples y dependen del tipo de organismo o célula en que se produzca. Dentro de los casos más estudiados se puede destacar la MDR en el tratamiento del cáncer [8], la epilepsia [9] y las enfermedades infecciosas [10], constituyendo uno de los mayores problemas que enfrenta la clínica hoy en día, pudiendo conducir al fracaso terapéutico y, en muchos casos, a la muerte del paciente.

La elevada frecuencia de aparición de MDR en la clínica parece estar en contradicción con el principio fundamental que establece que usar una combinación de fármacos no relacionados funcionalmente debería colaborar contra el desarrollo de una enfermedad resistente [11]. Se asume que la resistencia a un fármaco surge espontáneamente y está asociada a una dada tasa de mutación; la resistencia a dos fármacos no relacionados en la misma célula requeriría de dos eventos mutacionales independientes, generados con una frecuencia que depende del producto de las dos tasas de mutación. De esta suposición se desprende que la probabilidad de aparición de resistencia a tres o más fármacos independientes sería muy pequeña y no debería encontrarse frecuentemente. Sin embargo, las células de mamífero generan *in vitro* mutantes resistentes a múltiples fármacos a

una velocidad sorprendentemente elevada [12]. Este hecho sugiere que en ciertos casos el mecanismo de resistencia a múltiples fármacos no depende de eventos independientes, sino que está involucrado un mecanismo común.

Al investigar las causas de este fenómeno se caracterizó una forma específica de MDR: aquella mediada por una superfamilia de proteínas de membrana que han sido denominadas proteínas transportadoras de eflujo ATP dependientes o transportadores ABC (por sus siglas en inglés, *ATP-binding cassette*). Desde su descubrimiento, los transportadores ABC han sido arduamente estudiados en todo el mundo. Los mismos parecen encontrarse presentes en todas las especies vivas conocidas, tanto procariotas como eucariotas. Un dato importante para la estrategia implementada en este plan de tesis es que, además de encontrarse entre los principales responsables del fenómeno de MDR, los transportadores ABC participan en otras numerosas funciones fisiológicas [13].

No es para nada sorprendente que uno de los elementos fundamentales de la resistencia múltiple se encuentre localizado a nivel de la membrana plasmática, barrera que habitualmente el agente terapéutico debe superar para llegar a un blanco molecular intracelular o para atravesar epitelios y endotelios que funcionan como barrera para su acceso al órgano diana. Por supuesto que los transportadores ABC no evolucionaron para proteger a las células de la intervención médica; forman parte (junto con otros sistemas como el citocromo P450) de un antiguo y complejo mecanismo de defensa celular frente a las presiones impuestas externamente y potencialmente dañinas, que involucra el reconocimiento y la eliminación dependiente de energía de agentes químicos potencialmente tóxicos.

## ***1.2 Transportadores de eflujo ABC***

### ***1.2.1 Un poco de historia***

La primera evidencia que permitió vincular la MDR con el aumento en la tasa de eflujo de un fármaco surgió del trabajo de Keld Danø, quien demostró una reducción en la absorción de daunomicina en células del carcinoma ascítico

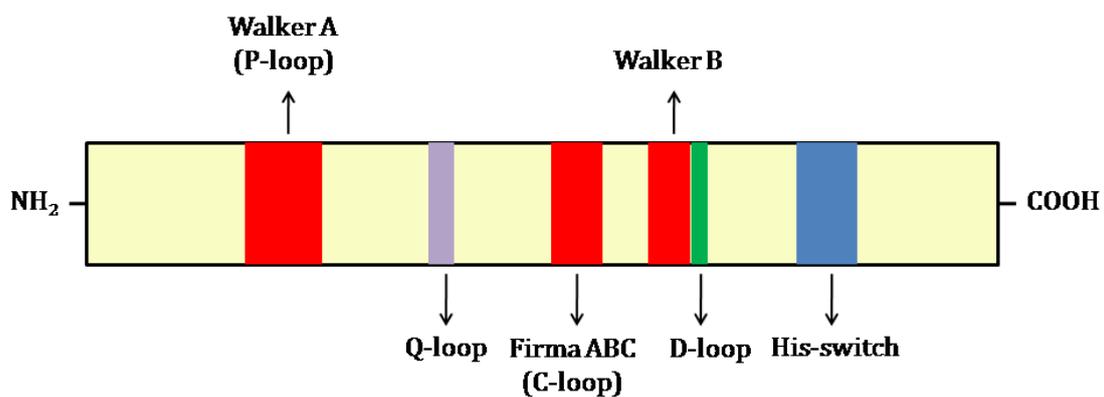
de Ehrlich, encontrando que esta reducción era dependiente de energía y presentaba competencia con alcaloides de la Vinca, fármacos frente a los que también evidenciaba resistencia cruzada [14], y por lo tanto propuso que la MDR estaba vinculada a un mecanismo de eflujo mediado por un transportador. Unos años más tarde, Victor Ling y col. [15] reportaron que clones de células de ovario de hámster chino seleccionados por su resistencia a colchicina mostraban resistencia cruzada frente a daunomicina y puromicina a la vez de una reducción en la captación de colchicina. Dado que el fenotipo de MDR se revertía cuando las células eran tratadas con fluidificantes de la membrana tales como detergentes suaves y anestésicos, se lo vinculó a una reducción de la permeabilidad de la membrana [16-17].

Numerosos interrogantes surgieron a partir de entonces, algunos de ellos se enfocaron en develar las bases genéticas del fenotipo MDR y su frecuencia. Gracias al análisis de fluctuación de Luria-Delbrück, originalmente desarrollado para demostrar la selección de mutantes que ocurren al azar en la aparición de fenotipos de resistencia bacteriana [18], se pudo confirmar que la aparición de células resistentes a múltiples fármacos se debía al aumento en la expresión de una glicoproteína de membrana de 170 kDa [19], que recibió el nombre de glicoproteína P (Pgp). El alto nivel de expresión de la Pgp en algunas líneas celulares permitió el estudio de la función de esta proteína en vesículas aisladas [20], y finalmente, la purificación y reconstitución de la Pgp pura en vesículas de membrana artificiales [21], proporcionando así una prueba definitiva de la capacidad de esta proteína para unir y transportar múltiples fármacos a través de la membrana plasmática.

El descubrimiento de otros transportadores ABC clínicamente significativos en mamíferos, comenzando con el regulador de la conductancia transmembrana de la fibrosis quística (CFTR o ABCC7) [22] y la MRP1 [23], cambió drásticamente la escena. A partir de la era post-genómica y hasta la fecha han sido identificados 49 transportadores ABC sólo en humanos, y un número mayor ha sido identificado en bacterias, parásitos, hongos y plantas [24-31].

### 1.2.2 Características estructurales

Como ya se mencionó, los transportadores ABC parecen estar presentes en todas las especies vivas conocidas ya que se han identificado en cada genoma que ha sido secuenciado hasta el momento. Llevan ese nombre debido a un dominio específico y altamente conservado denominado dominio ABC, constituido por una unidad proteica globular de 200 a 250 aminoácidos, cuya función principal es unir e hidrolizar trifosfato de adenosina (ATP). El dominio ABC, también llamado dominio de unión a nucleótidos o NBD (por el inglés, *nucleotide binding domain*), presenta varios motivos de secuencias conservadas (Figura 1.2). Estas secuencias, desde el-NH<sub>2</sub> al-COOH terminal serían: el Walker A (P-loop) una secuencia rica en glicina; una glutamina conservada (Q-loop), la firma ABC específica de familia (LSGGQ, también llamada C-loop), el motivo Walker B, una secuencia consenso (D-loop) y una histidina conservada (His-switch o H-loop). La firma ABC es una secuencia de diagnóstico para la superfamilia, ya que sólo está presente en las proteínas ABC, mientras que los motivos Walker A y B se encuentran en muchas otras proteínas que utilizan ATP [32].



**Figura 1.2.** Representación lineal del dominio ABC. Se muestran las secuencias conservadas. Las secuencias implicadas en la unión de nucleótidos están coloreadas en rojo.

Estructuralmente, los transportadores ABC funcionalmente activos presentan un mínimo de dos dominios ABC y dos dominios transmembrana (TMD). Los dominios ABC se encuentran localizados en el citoplasma y son encargados de unir e hidrolizar el ATP y por lo tanto de proporcionar la energía para los procesos que llevan a cabo estas proteínas. A diferencia de los dominios ABC, los dominios TMD se caracterizan por no presentar homología de secuencia entre los distintos miembros de la superfamilia; están constituidos por un número variable de  $\alpha$ -hélices transmembrana (típicamente entre 4 a 11) y son los responsables de conferir la especificidad de sustrato, de unirlo y translocarlo a través de la membrana [33-34]. Además, los dominios TMD dan lugar a bucles intracelulares (y en algunos casos extracelulares) denominados dominios o "loops" intracelulares (ICD) que conforman la interfaz entre los dominios ABC y los TMD y funcionan como hélices de acoplamiento encargadas de propagar la energía de la unión e hidrólisis de ATP para transportar los sustratos [35-36].

En algunos casos, los cuatro dominios principales están presentes en una única cadena polipeptídica, dando lugar a los llamados "transportadores completos". Por el contrario, los "medio-transportadores" poseen sólo una unidad ABC y una TMD y deben formar homodímeros o heterodímeros para generar un transportador ABC funcional [33].

Como se mencionó, el genoma humano codifica 49 proteínas ABC [37]. Dentro de la superfamilia existen grandes diferencias respecto al locus del gen, la secuencia de aminoácidos, la estructura y los sustratos que son transportados por cada proteína. Alineaciones de las secuencias de aminoácidos revelaron que estas proteínas se pueden agrupar en siete familias que van de la familia ABCA hasta la ABCG [32] (Tabla 1.1).

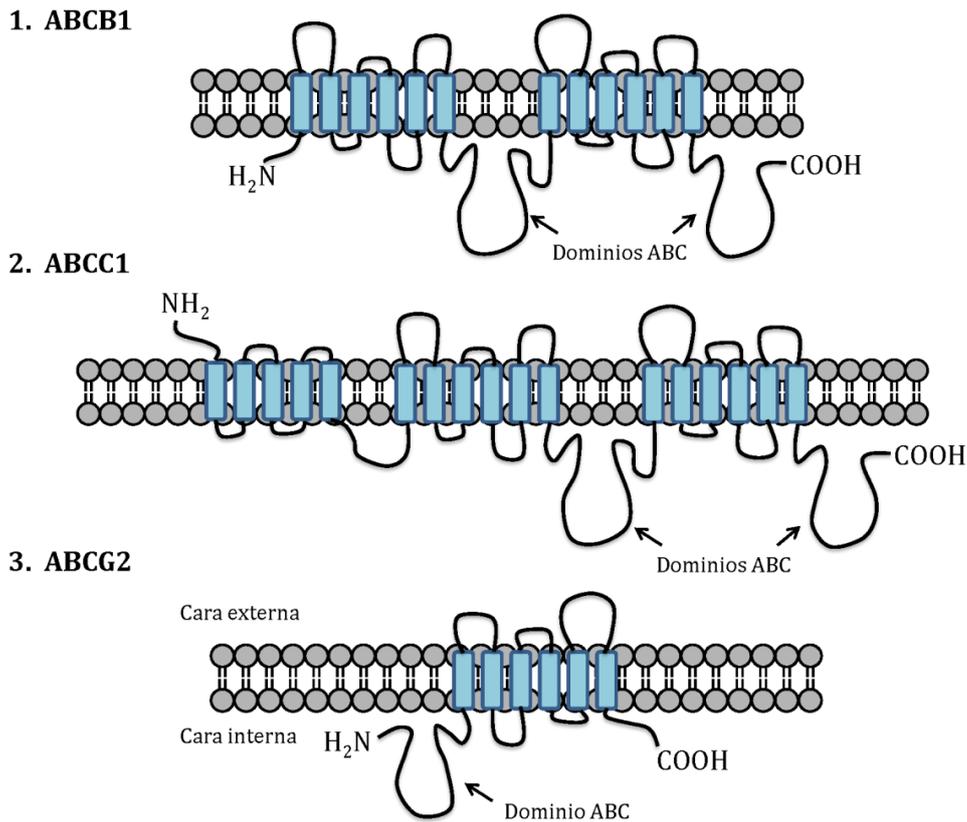
**Tabla 1.1.** Listado de los integrantes de las siete familias que conforman la superfamilia de transportadores ABC humanos y sus principales lugares de expresión [38-39].

<b>Familia</b>	<b>Símbolo</b>	<b>Sinónimos</b>	<b>Número de aminoácidos</b>	<b>Principales lugares de expresión</b>
ABCA (ABC1)	ABCA1	ABC1	2261	Ubicua
	ABCA2	ABC2	2436	Cerebro
	ABCA3	ABC3, ABCC	1704	Pulmón, cerebro, páncreas
	ABCA4	ABCR	2273	Células bastón de la retina
	ABCA5		1642	Músculo, corazón, testículos
	ABCA6		1617	Hígado
	ABCA7		2146	Bazo, timo
	ABCA8		1581	Ovario
	ABCA9		1624	Corazón
	ABCA10		1543	Músculo, corazón
	ABCA12		2595	Estómago
	ABCA13		5058	Bajo en todos los tejidos
	ABCB (MDR/TAP)	ABCB1	PGP, PGY1, MDR1	1280
ABCB2		TAP1	808	Ubicua, retículo endoplasmático
ABCB3		TAP2	703	Ubicua, retículo endoplasmático
ABCB4		PGY3, MDR3	1279	Hígado
ABCB5			812	Ubicua
ABCB6		MTABC3	842	Mitocondrias
ABCB7		ABC7	753	Mitocondrias
ABCB8		MABC1	718	Mitocondrias
ABCB9			766	Corazón, cerebro,

				lisosomas
	ABCB10	MTABC2	738	Mitocondrias
	ABCB11	SPGP, BSEP	1321	Hígado
ABCC (CFTR/MRP)	ABCC1	MRP1	1531	Pulmón, testículos, célula mononuclear de sangre periférica (PBMC)
	ABCC2	MRP2, cMOAT	1545	Hígado, intestino, riñón
	ABCC3	MRP3, cMOAT-2	1527	Pulmón, intestino, hígado
	ABCC4	MRP4, MOAT-B	1325	Próstata
	ABCC5	MRP5, MOAT-C	1437	Ubicua
	ABCC6	MRP6	1503	Riñón, hígado
	ABCC7	CFTR	1480	Pulmón, intestino, tejidos exocrinos
	ABCC8	SUR	1581	Páncreas
	ABCC9	SUR2	1549	Corazón, músculo
	ABCC10	MRP7	1464	Baja en todos los tejidos
	ABCC11	MRP8	1382	Baja en todos los tejidos
	ABCC12	MRP9	1359	Baja en todos los tejidos
	ABCC13	PRED6, MRP10	325	Hígado y bazo fetal, intestino.
ABCD (ALD)	ABCD1	ALD	745	Peroxisomas
	ABCD2	ALDL1, ALDR	740	Peroxisomas
	ABCD3	PXMP1, PMP70	659	Peroxisomas
	ABCD4	PMP69, P70R	606	Peroxisomas
ABCE (OABP)	ABCE1	OABP, RNS41	599	Ovario, testículos, bazo
ABCF (GCN20)	ABCF1	ABC50	845	Ubicua
	ABCF2		634	Ubicua
	ABCF3		709	Ubicua

ABCG (White)	ABCG1	ABC8, White	678	Ubicua
	ABCG2	BCRP, ABCP, MXR	655	Placenta, intestino, hígado, cerebro
	ABCG4	White2	646	Hígado
	ABCG5	White3, Steroline 1	651	Hígado, intestino
	ABCG8	Steroline 2	673	Hígado, intestino

Se ha observado que cada familia se caracteriza por patrones de topología de membrana típicos (Figura 1.3). Brevemente, los miembros de la familia ABCA son transportadores completos con una disposición de dominios de tipo TMD1-ABC1-TMD2-ABC2. Los TMD de esta familia contienen bucles extracelulares muy grandes con numerosos sitios de glicosilación. La familia ABCB está conformada por cuatro transportadores completos (incluyendo a Pgp) con una disposición de dominios de tipo TMD1-ABC1-TMD2-ABC2, y por siete medio-transportadores [32]. Los miembros de la familia ABCC se pueden dividir en dos grupos: los "cortos" (CFTR, MRP4, MRP5, MRP8, MRP9 y MRP10) que presentan el típico arreglo TMD-ABC de los 4 dominios, y los "largos" (MRP1, MRP2, MRP3, MRP6, MRP7, ABCC8 y ABCC9) que presentan además de los 4 dominios una región TMD adicional en el extremo NH<sub>2</sub>-terminal compuesta por 200 aminoácidos [40]. La familia ABCD incluye cuatro medio-transportadores con el típico arreglo TMD-ABC. Los miembros de las familias ABCE y ABCF no están involucrados en los procesos de transporte de membrana y carecen de dominios TMD. Los cinco medio-transportadores en la familia ABCG muestran una disposición de dominios inversa (ABC-TMD) [32].



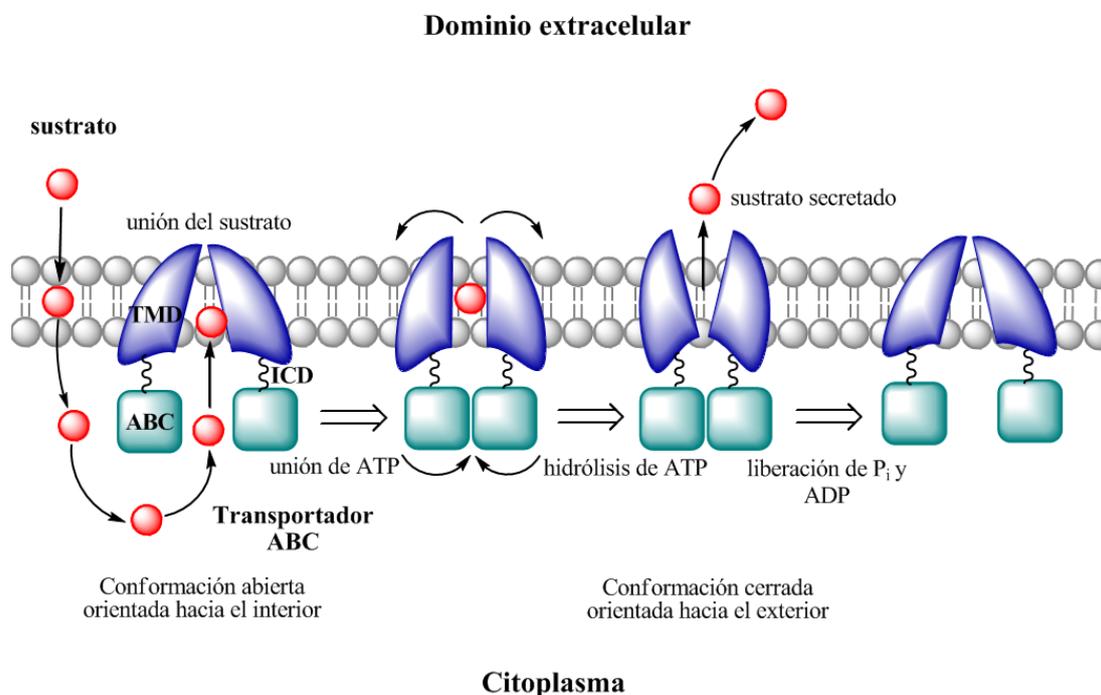
**Figura 1.3.** Representación de los tres patrones de topología de membrana más típicos que presentan los transportadores ABC. **1.** ABCB1 (P-gp) contiene 12 dominios transmembrana (TMD) y dos regiones de unión a ATP (ABC). **2.** ABCC1 (MRP1) contiene 2 regiones de unión a ATP (ABC) y un dominio adicional compuesto por 5 segmentos transmembrana en el extremo amino-terminal, dando un total de 17 segmentos TMD. **3.** ABCG2 (BCRP, proteína de resistencia del cáncer de mama) es un medio-transportador y contiene seis segmentos TMD y una región de unión a ATP (ABC) en el lado amino-terminal del TMD (disposición de dominios inversa). *Adaptada de [41].*

En un primer momento, las estructuras ABC fueron obtenidas como monómeros, y los autores sugirieron un dímero, en el que las dos unidades ABC fueron colocados en una orientación “*back-to-back*” o espalda con espalda [42]. Sin embargo, este ensamblado representa una interacción energéticamente desfavorable de los dos monómeros y sugiere dos sitios de unión a nucleótidos (catalíticos) altamente expuestos, lo cual es difícil de conciliar con la función altamente regulada de los sitios activos. Finalmente se ha establecido que las dos subunidades ABC que interactúan funcionalmente dimerizan en una orientación “*head-to-tail*” o cabeza-cola. Los sitios activos de los dos dominios ABC se

complementan entre sí, formando dos centros catalíticos compuestos. La secuencia Walker A de una subunidad y la firma ABC de la subunidad opuesta están involucrados en la formación de cada uno de los dos sitios de unión a ATP [43-45].

### 1.2.3 Función

Los transportadores ABC humanos, y eucariotas en general, utilizan la energía proveniente de la hidrólisis del ATP intracelular para exportar compuestos desde el citoplasma hacia el exterior de la célula a través de la membrana plasmática [46], o bien para mover moléculas en orgánulos intracelulares, como por ejemplo a través del retículo endoplasmático, las mitocondrias o los peroxisomas (Figura 1.4) [33]. En cualquier caso, el sentido de transporte en eucariotas es siempre desde el interior de la célula u orgánulo hacia el exterior de éstos [39, 47].



**Figura 1.4.** Esquema del funcionamiento de un transportador ABC. ATP: trifosfato de adenosina, ADP: difosfato de adenosina,  $P_i$ : fosfato inorgánico, TMD: Dominio transmembrana, ABC: Dominio de unión a nucleótidos, ICD: Dominio intracelular.

En los mamíferos, los transportadores ABC se expresan de forma ubicua en todo el organismo (Tabla 1.1), pero su expresión es predominante en ciertos órganos como el hígado, el intestino y los riñones, y en tejidos tales como la barrera hematoencefálica, la barrera hematotesticular y la placenta [38]. Están involucrados en el eflujo tanto de moléculas endógenas como de compuestos xenobióticos, y por lo tanto están asociados a un amplio espectro de funciones fisiológicas, incluyendo la excreción de toxinas en el hígado, los riñones y el tracto gastrointestinal, el metabolismo lipídico y la presentación de antígenos, entre otras [48]. Su importancia en la modulación de los procesos de absorción, distribución, metabolismo, secreción y excreción es cada vez más clara [49].

Cabe destacar que existen varias proteínas ABC humanas asociadas a la membrana con funciones predominantes de tipo canal o incluso de tipo receptor. Mientras que los transportadores ABC hidrolizan ATP en estrecho acoplamiento con el movimiento transmembrana de una molécula sustrato, los canales y los receptores utilizan los dominios de unión a ATP sobre todo para la regulación de la apertura y/o cierre de vías, lo que permite el paso de iones o la transmisión de información a través de la membrana. Entre los transportadores ABC humanos que llevan a cabo tales funciones y se encuentran bien caracterizados podemos mencionar el ABCC7/CFTR (canal de ion cloruro), ABCC8/SUR1 y ABCC9/SUR2, receptores de sulfonilureas que actúan como subunidades reguladoras de los canales de potasio dependientes de ATP en las células beta productoras de insulina del páncreas y en el corazón, respectivamente [32].

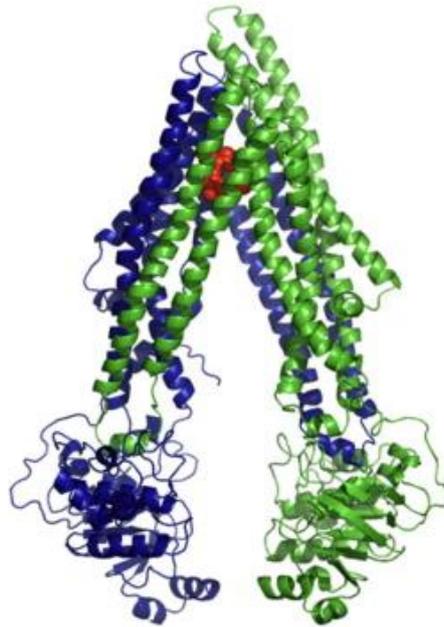
Queda de manifiesto entonces que la especificidad de sustrato de estas proteínas resulta tan amplia como su distribución tisular y su función fisiológica. Si bien cada transportador presenta un espectro de sustratos particular, en algunos casos la especificidad de sustrato se solapa entre dos o más transportadores de la superfamilia, resultando que una molécula puede ser sustrato de más de un transportador al mismo tiempo [50]. Los sustratos difieren enormemente en tamaño y naturaleza química y dentro de este enorme universo se pueden mencionar aminoácidos, ácidos grasos, colesterol y sus derivados, azúcares, vitaminas, péptidos y proteínas pequeñas, lípidos, aniones inorgánicos, iones metálicos, intermediarios metabólicos, hormonas y un gran número de

compuestos hidrofóbicos, entre ellos numerosos fármacos y sus metabolitos [33, 37-39].

Como son esenciales para muchos procesos celulares, las mutaciones en al menos 18 genes ABC causan o contribuyen a varios trastornos genéticos, entre ellos: fibrosis quística (CFTR/ABCC7); enfermedad de Tangier (ABCA1); enfermedad de Stargardt, retinitis pigmentaria y degeneración macular relacionada con la edad (ABCA4); espondilitis anquilosante, diabetes tipo 2, enfermedad celíaca y anemia sideroblástica ligada al cromosoma X con ataxia (varios miembros de la familia ABCB); colestasis intrahepática familiar progresiva (ABCB11); síndrome de Dubin-Johnson (ABCC2); pseudoxantoma elástico (ABCC6); adrenoleucodistrofia ligada al cromosoma X (ABCD1 y ABCD2); algunas formas del síndrome de Zellweger (ABCD3 y ABCD2) y; sitosterolemia (ABDG5 y ABCG8) [33, 38-39, 47].

#### *1.2.4 Mecanismo de acción*

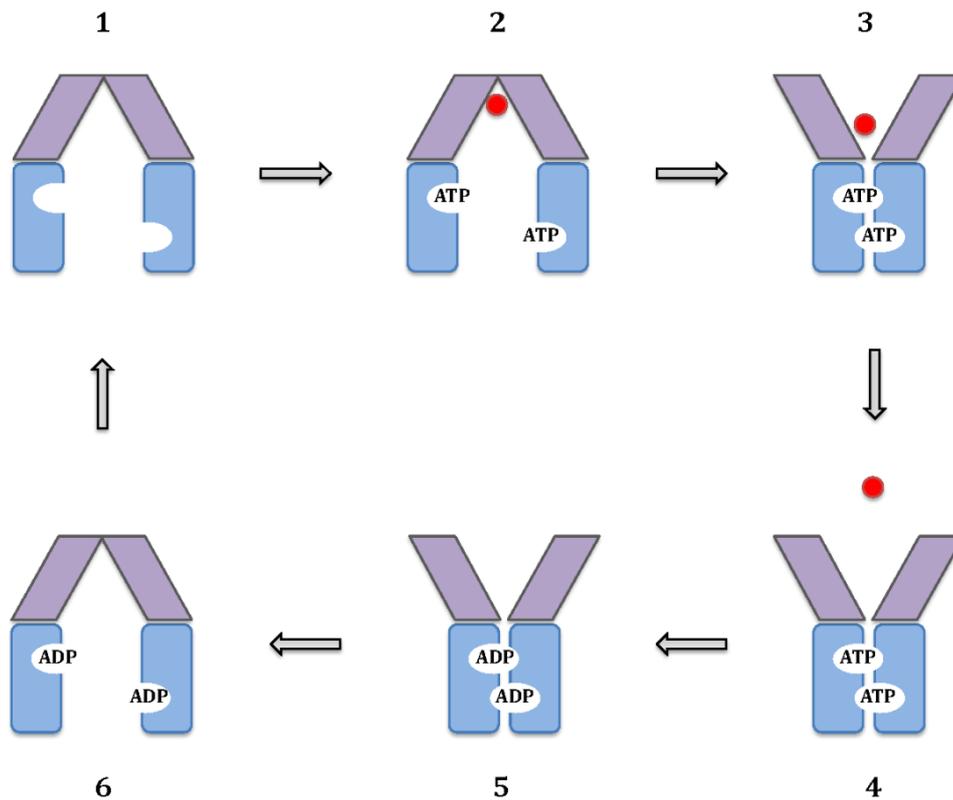
En base a más de una docena de estructuras cristalográficas resueltas (ver, por ejemplo, Figura 1.5) y a gran cantidad de datos bioquímicos y biofísicos, se han propuesto una serie de modelos para explicar el mecanismo por el cual los transportadores ABC son capaces de translocar sustratos a través de la membrana plasmática [34-35, 51].



**Figura 1.5.** Estructura 3D del transportador ABCB1 de ratón obtenida por cristalografía de rayos X. Se encuentra en la conformación abierta orientada hacia el interior celular, con los dominios ABC libres de nucleótidos a 30 Å de distancia. El inhibidor se indica en color rojo. Cada dominio TMD-ABC se representa en azul y verde, respectivamente.

El paradigma imperante es el denominado “modelo de interruptor de ATP” (o “*ATP Switch Model*”) [34-35, 52-54], según el cual en el estado de reposo, el transportador está en una conformación abierta orientada hacia el interior celular con el sitio de unión a sustratos de “alta afinidad” expuesto hacia el lado citoplasmático y la cara interior de la membrana (Figura 1.5). Se produce la unión del sustrato dentro de los dominios TMD y, tras la unión de dos moléculas de ATP, los dominios ABC dimerizan formando un dímero simétrico cerrado. Este cambio conformacional de los dominios ABC de abierto a cerrado provoca un cambio conformacional en los dominios TMD que lleva al transportador a una conformación orientada hacia el exterior. Esta nueva conformación permite la extrusión del sustrato y su posterior liberación hacia el exterior celular, que ocurre probablemente como consecuencia de la disminución de la afinidad de unión causada por cambios en los contactos de residuos específicos entre la proteína y el sustrato en un sitio de “baja afinidad”. La hidrólisis secuencial de las dos moléculas de ATP y posterior liberación de ADP/Pi provoca la disociación del dímero ABC, lo

que induce cambios en los TMD que vuelven a estar orientados hacia el interior, completando así el ciclo de transporte (Figura 1.6). *In vivo*, debido a la alta concentración de ATP intracelular, el ATP podría estar siempre unido a los dominios ABC y por lo tanto, la transición del transportador a la conformación orientada hacia el exterior se desencadena por la unión del sustrato.



**Figura 1.6.** Esquema del modelo de interruptor de ATP ("ATP Switch Model"). Los dominios TMD se representan en violeta, los ABC como rectángulos celestes y el sustrato como un punto rojo. **1.** Dominios ABC separados y libres de nucleótidos en una conformación abierta orientada hacia el interior celular. **2.** Se une el sustrato a los dominios TMD y dos moléculas de ATP a los dominios ABC. **3.** Los dominios ABC dimerizan con las dos moléculas de ATP unidas en su interfaz. La conformación de los dominios TMD cambia a una orientada hacia el exterior y el sustrato se expone a los sitios de baja afinidad. **4.** Se libera el sustrato y se hidroliza el ATP. **5.** En el dímero ABC permanecen unidas dos moléculas de ADP. Puede implicar una conformación semi-abierta de los dominios ABC no representada. **6.** Las moléculas de ADP (que luego se liberan) inducen la separación del dímero ABC, provocando la reorientación de los dominios TMD a la conformación orientada hacia el interior, completando el ciclo. *Esquema adaptado de [34].*

Este modelo predice la hidrólisis de ATP en ambos sitios catalíticos por ciclo. Sin embargo, existe evidencia sustancial de que los dominios ABC no se separan mucho (si es que lo hacen), por lo que la separación física de los ABC observada en las estructuras cristalográficas puede ser un artefacto producto del mismo proceso cristalográfico (Figura 1.5) [34]. Cabe la posibilidad que la hidrólisis de ATP ocurra en un solo sitio, siendo suficiente para la restauración de la conformación orientada hacia el interior. Esta idea se encuentra en concordancia con un modelo alternativo denominado "modelo de contacto constante" (o "*Constant Contact Model*") [35, 51, 55-56] que también cuenta con una amplia aceptación en buena parte de la comunidad científica. Este modelo alternativo propone que los dominios ABC no se disocian completamente sino que permanecen en contacto durante todo el ciclo catalítico, y la hidrólisis de ATP se produce alternativamente en cada uno de los dos sitios activos, siendo cada sitio capaz de abrir y de intercambiar nucleótidos mientras que el otro sitio permanece cerrado.

### **1.3 Transportadores MDR-ABC**

La investigación de los transportadores ABC ha centrado su atención en aquellos que han sido vinculados con el fenómeno de MDR y por lo tanto denominados transportadores MDR-ABC. De los 49 transportadores humanos, sólo la proteína ABCB1 (Pgp, MDR1), la proteína ABCG2 (BCRP, MXR) y varios miembros de la familia ABCC (MRP) califican sin dudas como proteínas MDR-ABC y su participación en este fenómeno ha sido largamente documentada, principalmente en el cáncer, la epilepsia y las enfermedades infecciosas [57-58].

De aquí en adelante, el análisis se centrará en las causas propuestas -hasta el momento- de la MDR en la epilepsia refractaria, especialmente al papel que desempeñan los transportadores MDR-ABC y fundamentalmente al rol protagónico que en los últimos años ha tomado uno de ellos: la proteína de resistencia del cáncer de mama (ABCG2/BCRP).

---

## ***Referencias Capítulo I***

- [1] Sheps JA, Ling V. Introduction: What is Multidrug Resistance? In: Boumendjel A, Boutonnat J, Robert J, editors. ABC Transporters and Multidrug Resistance. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2009.
- [2] Nassau E, Hamilton GM. Multiple drug resistance and virulence of mycobacterium tuberculosis; a preliminary communication. *Tubercle*. 1955;36(9):281-2.
- [3] Burchenal JH, Coley V, Purple JR, Bucholz E, Lyman MS, Kreis W. Studies on the Mechanisms of Action of Various Phthalanilide Derivatives by Cross-Resistance and Tissue Culture. *Cancer Res*. 1963;23:1364-74.
- [4] Kessel D, Botterill V, Wodinsky I. Uptake and retention of daunomycin by mouse leukemic cells as factors in drug response. *Cancer Res*. 1968;28(5):938-41.
- [5] Kessel D, Wodinsky I. Uptake in vivo and in vitro of actinomycin D by mouse leukemias as factors in survival. *Biochem Pharmacol*. 1968;17(1):161-4.
- [6] Biedler JL, Riehm H. Cellular resistance to actinomycin D in Chinese hamster cells in vitro: cross-resistance, radioautographic, and cytogenetic studies. *Cancer Res*. 1970;30(4):1174-84.
- [7] Riehm H, Biedler JL. Cellular resistance to daunomycin in Chinese hamster cells in vitro. *Cancer Res*. 1971;31(4):409-12.
- [8] Wu Q, Yang Z, Nie Y, Shi Y, Fan D. Multi-drug resistance in cancer chemotherapeutics: mechanisms and lab approaches. *Cancer Lett*. 2014;347(2):159-66.
- [9] Xiong J, Mao DA, Liu LQ. Research Progress on the Role of ABC Transporters in the Drug Resistance Mechanism of Intractable Epilepsy. *Biomed Res Int*. 2015;2015:194541.
- [10] Tanwar J, Das S, Fatima Z, Hameed S. Multidrug resistance: an emerging crisis. *Interdiscip Perspect Infect Dis*. 2014;2014:541340.
- [11] Goldie JH, Coldman AJ. A mathematic model for relating the drug sensitivity of tumors to their spontaneous mutation rate. *Cancer Treat Rep*. 1979;63(11-12):1727-33.
- [12] Thompson LH, Baker RM. Isolation of mutants of cultured mammalian cells. *Methods Cell Biol*. 1973;6:209-81.

- [13] Holland IB, Cole SPC, Kuchler K, Higgins CF. ABC proteins: from bacteria to man. London: Academic Press; 2003.
- [14] Dano K. Active outward transport of daunomycin in resistant Ehrlich ascites tumor cells. *Biochim Biophys Acta*. 1973;323(3):466-83.
- [15] Ling V, Thompson LH. Reduced permeability in CHO cells as a mechanism of resistance to colchicine. *J Cell Physiol*. 1974;83(1):103-16.
- [16] Ling V. Drug resistance and membrane alteration in mutants of mammalian cells. *Can J Genet Cytol*. 1975;17(4):503-15.
- [17] Carlsen SA, Till JE, Ling V. Modulation of membrane drug permeability in Chinese hamster ovary cells. *Biochim Biophys Acta*. 1976;455(3):900-12.
- [18] Luria SE, Delbruck M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics*. 1943;28(6):491-511.
- [19] Riordan JR, Ling V. Purification of P-glycoprotein from plasma membrane vesicles of Chinese hamster ovary cell mutants with reduced colchicine permeability. *J Biol Chem*. 1979;254(24):12701-5.
- [20] Horio M, Gottesman MM, Pastan I. ATP-dependent transport of vinblastine in vesicles from human multidrug-resistant cells. *Proc Natl Acad Sci U S A*. 1988;85(10):3580-4.
- [21] Shapiro AB, Ling V. Reconstitution of drug transport by purified P-glycoprotein. *J Biol Chem*. 1995;270(27):16167-75.
- [22] Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*. 1989;245(4922):1066-73.
- [23] Cole SP, Bhardwaj G, Gerlach JH, Mackie JE, Grant CE, Almquist KC, et al. Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science*. 1992;258(5088):1650-4.
- [24] Giacomini KM, Huang SM, Tweedie DJ, Benet LZ, Brouwer KL, Chu X, et al. Membrane transporters in drug development. *Nat Rev Drug Discov*. 2010;9(3):215-36.
- [25] Liu H, Xie J. Comparative genomics of Mycobacterium tuberculosis drug efflux pumps and their transcriptional regulators. *Crit Rev Eukaryot Gene Expr*. 2014;24(2):163-80.

- [26] Dassa E. Phylogenetic and functional classification of ABC (ATP-binding cassette) systems. In: Holland IB, Cole SPC, Kuchler K, Higgins CF, editors. ABC proteins: from bacteria to man. London: Academic Press; 2003. p. 3-35.
- [27] Ardelli BF. Transport proteins of the ABC systems superfamily and their role in drug action and resistance in nematodes. *Parasitol Int.* 2013;62(6):639-46.
- [28] Greenberg RM. ABC multidrug transporters in schistosomes and other parasitic flatworms. *Parasitol Int.* 2013;62(6):647-53.
- [29] Paul S, Moye-Rowley WS. Multidrug resistance in fungi: regulation of transporter-encoding gene expression. *Front Physiol.* 2014;5:143.
- [30] Szewczak A, Ziomkiewicz I, Jasiński M. Hiring cell gatekeepers – ABC transporters in plant biotechnology. *Biotechnologia.* 2011;92(2):132-9.
- [31] Knöllner AS, Murphy AS. ABC Transporters and Their Function at the Plasma Membrane. In: Murphy AS, Schulz B, Peer W, editors. *The Plant Plasma Membrane*: Springer; 2011. p. 353-77.
- [32] Sarkadi B, Homolya L, Szakacs G, Varadi A. Human multidrug resistance ABCB and ABCG transporters: participation in a chemoimmunity defense system. *Physiol Rev.* 2006;86(4):1179-236.
- [33] Licht A, Schneider E. ATP binding cassette systems: structures, mechanisms, and functions. *Central European Journal of Biology.* 2011;6(5):785-801.
- [34] Jones PM, George AM. A reciprocating twin-channel model for ABC transporters. *Q Rev Biophys.* 2014;47(3):189-220.
- [35] George AM, Jones PM. Perspectives on the structure-function of ABC transporters: the Switch and Constant Contact models. *Prog Biophys Mol Biol.* 2012;109(3):95-107.
- [36] Dawson RJ, Locher KP. Structure of a bacterial multidrug ABC transporter. *Nature.* 2006;443(7108):180-5.
- [37] Kathawala RJ, Gupta P, Ashby CR, Jr., Chen ZS. The modulation of ABC transporter-mediated multidrug resistance in cancer: a review of the past decade. *Drug Resist Updat.* 2015;18:1-17.
- [38] Vasiliou V, Vasiliou K, Nebert DW. Human ATP-binding cassette (ABC) transporter family. *Hum Genomics.* 2009;3(3):281-90.
- [39] Dean M, Hamon Y, Chimini G. The human ATP-binding cassette (ABC) transporter superfamily. *J Lipid Res.* 2001;42(7):1007-17.

- [40] Deeley RG, Westlake C, Cole SP. Transmembrane transport of endo- and xenobiotics by mammalian ATP-binding cassette multidrug resistance proteins. *Physiol Rev.* 2006;86(3):849-99.
- [41] Joyce H, McCann A, Clynes M, Larkin A. Influence of multidrug resistance and drug transport proteins on chemotherapy drug metabolism. *Expert Opin Drug Metab Toxicol.* 2015;11(5):795-809.
- [42] Hung LW, Wang IX, Nikaido K, Liu PQ, Ames GF, Kim SH. Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature.* 1998;396(6712):703-7.
- [43] Smith PC, Karpowich N, Millen L, Moody JE, Rosen J, Thomas PJ, et al. ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell.* 2002;10(1):139-49.
- [44] Zaitseva J, Jenewein S, Jumpertz T, Holland IB, Schmitt L. H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *EMBO J.* 2005;24(11):1901-10.
- [45] Hollenstein K, Dawson RJ, Locher KP. Structure and mechanism of ABC transporter proteins. *Curr Opin Struct Biol.* 2007;17(4):412-8.
- [46] Rees DC, Johnson E, Lewinson O. ABC transporters: the power to change. *Nat Rev Mol Cell Biol.* 2009;10(3):218-27.
- [47] Dean M, Rzhetsky A, Allikmets R. The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res.* 2001;11(7):1156-66.
- [48] Holland IB. ABC transporters, mechanisms and biology: an overview. *Essays Biochem.* 2011;50(1):1-17.
- [49] Szakacs G, Varadi A, Ozvegy-Laczka C, Sarkadi B. The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). *Drug Discov Today.* 2008;13(9-10):379-93.
- [50] Marquez B, Van Bambeke F. ABC multidrug transporters: target for modulation of drug pharmacokinetics and drug-drug interactions. *Curr Drug Targets.* 2011;12(5):600-20.
- [51] Jones PM, George AM. Mechanism of the ABC transporter ATPase domains: catalytic models and the biochemical and biophysical record. *Crit Rev Biochem Mol Biol.* 2013;48(1):39-50.
- [52] Higgins CF, Linton KJ. The ATP switch model for ABC transporters. *Nat Struct Mol Biol.* 2004;11(10):918-26.

- [53] Linton KJ, Higgins CF. Structure and function of ABC transporters: the ATP switch provides flexible control. *Pflugers Arch.* 2007;453(5):555-67.
- [54] van der Does C, Tampe R. How do ABC transporters drive transport? *Biol Chem.* 2004;385(10):927-33.
- [55] Jones PM, George AM. Opening of the ADP-bound active site in the ABC transporter ATPase dimer: evidence for a constant contact, alternating sites model for the catalytic cycle. *Proteins.* 2009;75(2):387-96.
- [56] Siarheyeva A, Liu R, Sharom FJ. Characterization of an asymmetric occluded state of P-glycoprotein with two bound nucleotides: implications for catalysis. *J Biol Chem.* 2010;285(10):7575-86.
- [57] Choi YH, Yu AM. ABC transporters in multidrug resistance and pharmacokinetics, and strategies for drug development. *Curr Pharm Des.* 2014;20(5):793-807.
- [58] Gottesman MM, Ling V. The molecular basis of multidrug resistance in cancer: the early years of P-glycoprotein research. *FEBS Lett.* 2006;580(4):998-1009.



## Capítulo II

# Epilepsia refractaria

---



## 2.1 Definiciones

La epilepsia es el trastorno cerebral crónico más frecuente, afectando a más de 50 millones de personas en todo el mundo, y representa el segundo trastorno neurológico más común (luego de los accidentes cerebro-vasculares) [1-2]. La prevalencia de este desorden es mayor en países en vías de desarrollo posiblemente debido al mayor riesgo de sufrir afecciones que puedan producir daño cerebral permanente [2-3]. Se trata de un desorden neurológico heterogéneo y complejo, debido a su origen y sintomatología multifactorial, que afecta a personas de todas las edades [4]. Las manifestaciones sintomáticas de la epilepsia se denominan convulsiones, ataques o crisis epilépticas, las cuales son muy variadas dependiendo de la región del cerebro en que se originan y son el resultado de las descargas eléctricas excesivas en un grupo de células cerebrales. Habitualmente, las convulsiones se manifiestan como episodios de movimientos involuntarios que pueden implicar una parte del cuerpo (parcial) o todo el cuerpo (generalizada), y en ocasiones se acompañan de pérdida de la conciencia y el control de la función intestinal o de la vejiga. Pueden variar desde breves lapsos de pérdida de atención o tirones musculares a convulsiones severas y prolongadas, y también pueden variar en frecuencia, desde menos de una por año a varias por día. [2].

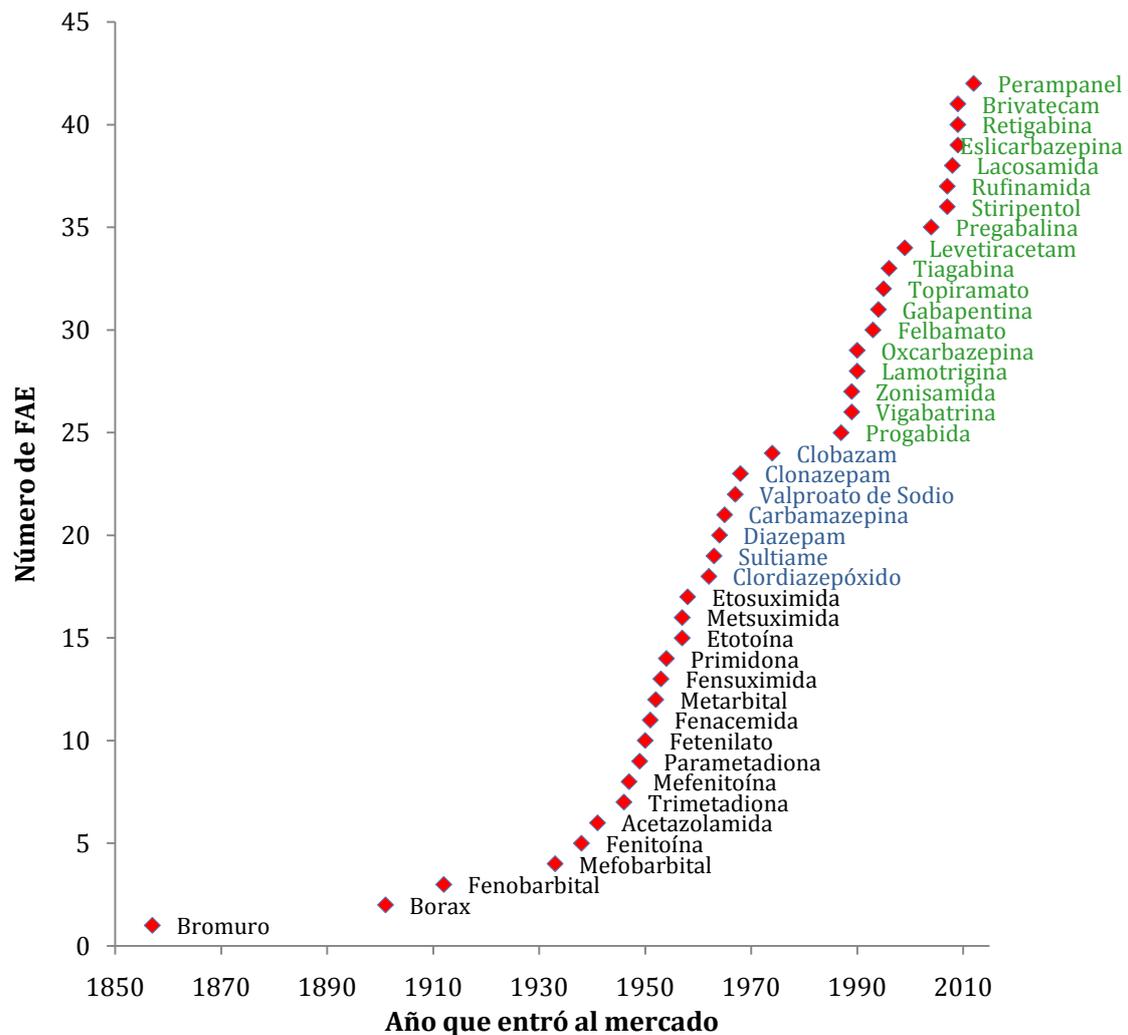
El tipo más común de epilepsia, que afecta a 6 de cada 10 personas con este desorden, se denomina epilepsia idiopática y no tiene ninguna causa identificable. La epilepsia con una causa conocida se conoce como epilepsia secundaria o epilepsia sintomática. Las causas de la epilepsia secundaria podrían ser: daño cerebral por lesiones prenatales o perinatales (por ejemplo, una pérdida de oxígeno o trauma durante el parto), anomalías congénitas o enfermedades genéticas con malformaciones cerebrales asociadas, un traumatismo en el encéfalo, un derrame cerebral que limita la cantidad de oxígeno en el cerebro, una infección del cerebro (meningitis, encefalitis, neurocisticercosis), ciertos síndromes genéticos, un tumor cerebral, entre otras [2].

Para tratar este trastorno es necesario restablecer el balance dinámico entre la neurotransmisión inhibitoria y excitatoria, y para ello la farmacoterapia es el tratamiento de elección logrando éxito en el control de la epilepsia (es decir,

logrando un estado libre de convulsiones sostenido en el tiempo) en alrededor del 70% de los pacientes [2]. El 30% restante sufre de epilepsia resistente al tratamiento, refractaria o intratable [5], en la cual no se logra alcanzar un estado libre de convulsiones sostenido en el tiempo con al menos dos regímenes de fármacos antiepilépticos (FAE) correctamente seleccionados, bien tolerados y de uso habitual [6-7]. En los pacientes que no responden al primer fármaco suministrado, la probabilidad de respuesta disminuye con el número de otros FAE probados sin éxito [5, 8].

Desde la introducción del bromuro en 1857, el desarrollo de nuevos FAE ha tenido un aumento exponencial (Figura 2.1); la introducción de más de 15 nuevos FAE desde 1990 hasta el momento [9] ha permitido mejorar la tolerancia a los tratamientos, reducir los efectos adversos, lograr una menor interacción con otros fármacos y una mejor farmacocinética. Sin embargo, a pesar de todo este nuevo arsenal terapéutico disponible el número de pacientes refractarios no ha cambiado significativamente [9]. Este hecho ha dado lugar a una disminución en el interés de la industria en el desarrollo de nuevos compuestos para el tratamiento de la epilepsia [10].

Lamentablemente las consecuencias de la epilepsia no controlada pueden ser graves incluyendo el aumento del riesgo de muerte (lo que se asocia a una menor expectativa de vida), lesiones corporales y trastornos de índole somática, deterioro neuropsicológico y psiquiátrico, y discapacidad social [11-13].



**Figura 2.1.** Evolución de la introducción al mercado de los FAE más relevantes entre 1853 a 2009. Se grafica el año de la primera licencia o la primera mención del uso clínico en un país de Europa, Estados Unidos o Japón. En negro se indican los FAE de primera generación, en azul los de segunda y en verde los de tercera generación [14-15].

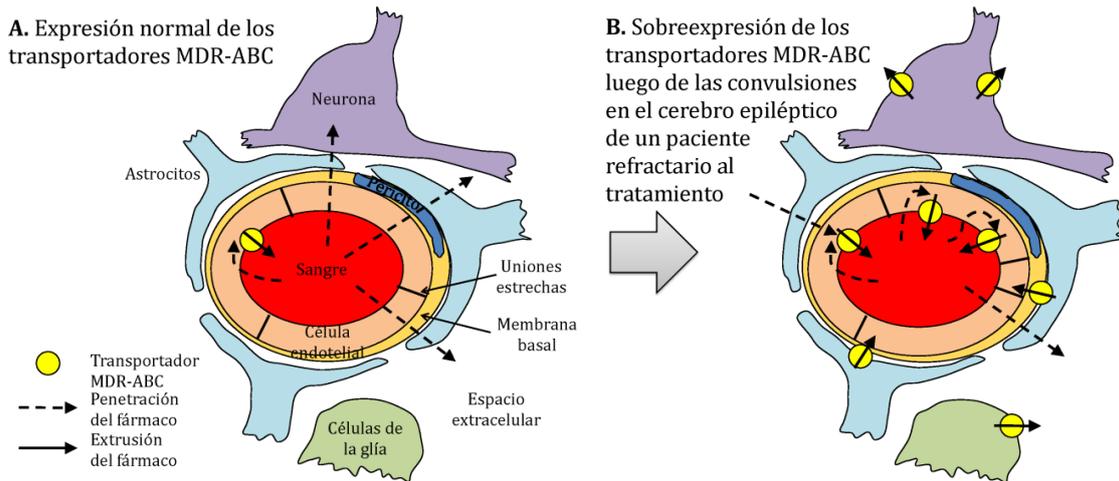
## 2.2 Hipótesis de la epilepsia refractaria

La resistencia a los FAE puede existir desde la primera convulsión o se puede desarrollar posteriormente como resultado del proceso del propio trastorno, por lo que se piensa que puede ser consecuencia de alteraciones tanto constitutivas (intrínsecas) como inducidas o fisiopatológicas, por ejemplo, como consecuencia de convulsiones recurrentes o de la progresión del cuadro [16]. Los mecanismos biológicos responsables de la epilepsia refractaria aún no han sido completamente

dilucidados [10], por lo que existen actualmente al menos cinco hipótesis que tratan de explicar la naturaleza de este fenómeno:

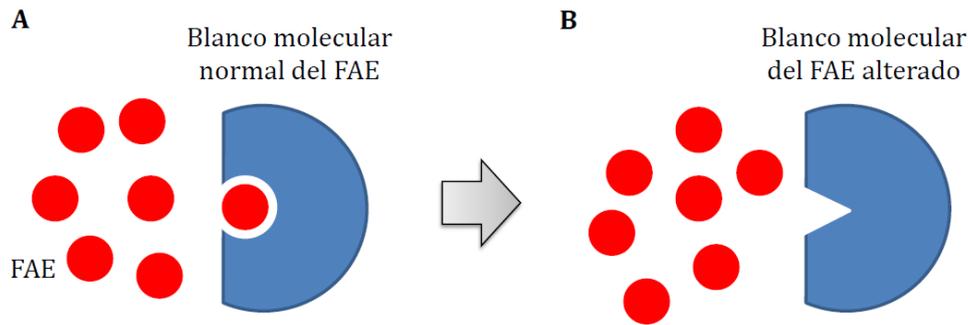
1. hipótesis de los transportadores [17-18],
2. hipótesis del blanco molecular [18-19],
3. hipótesis de las redes neuronales [20],
4. hipótesis de la severidad intrínseca [21],
5. hipótesis de la variante genética [22].

La hipótesis de los transportadores y la del blanco molecular fueron las primeras en ser postuladas y han sido exploradas experimentalmente con mayor intensidad, por lo que acumulan gran cantidad de evidencia y son las más aceptadas. En este sentido, se ha reunido un importante volumen de evidencia clínica y experimental (revisada en mayor profundidad en la siguiente sección) que sustenta a la hipótesis de los transportadores, la cual sostiene que la epilepsia refractaria puede ser una consecuencia de una sobreexpresión e hiperactividad local de los transportadores MDR-ABC en la barrera hemato-encefálica (BHE) y/o los focos epilépticos (Figura 2.2), los cuales limitarían la penetración de los FAE y conducirían a niveles insuficientes de los fármacos en el tejido cerebral epileptogénico [10].



**Figura 2.2.** Representación esquemática de la hipótesis de los transportadores. **A.** Nivel de expresión normal de los transportadores MDR-ABC en la unidad neurovascular y, **B.** Sobreexpresión luego de las convulsiones en los pacientes refractarios al tratamiento. La BHE está formada por las células endoteliales de los capilares, que están unidas por uniones estrechas. El endotelio capilar está rodeado por una membrana basal y pies perivasculares de los astrocitos, que contribuyen a la función de la BHE. Los capilares cerebrales actúan de una manera pasiva para restringir la penetración de sustancias hidrofílicas, ionizadas (polares) y/o de alto peso molecular, pero los fármacos altamente lipofílicos (como la mayoría de los FAE) penetran fácilmente a través de la BHE por difusión pasiva (A). Como mecanismo de defensa activa contra de las sustancias lipofílicas, se expresan normalmente a nivel de la BHE los transportadores de eflujo ABC, que se encuentran en el lado apical (luminal) de la membrana celular de las células endoteliales y actúan como bombas de eflujo activas, transfiriendo parte de los fármacos sustrato que han ingresado a las células endoteliales de nuevo a la sangre, lo que limita la penetración de muchos fármacos lipofílicos en el parénquima cerebral. Debido a que los FAE son altamente lipofílicos y pequeños, la expresión constitutiva de los transportadores ABC no restringe críticamente su captación cerebral. Como se ilustra en B, en el tejido cerebral epileptogénico de pacientes con epilepsia refractaria al tratamiento, se produce una sobreexpresión de estos transportadores ABC en toda la unidad neurovascular reduciendo críticamente la captación cerebral y aumentando el eflujo de los FAE sustrato de tales transportadores. *Adaptado de [23].*

Por su parte, la hipótesis del blanco molecular propone que la reducción de la sensibilidad a los FAE podría estar vinculada a modificaciones adquiridas en la estructura y/o funcionalidad de los blancos moleculares de los FAE (Figura 2.3) como son los canales iónicos (por ejemplo el canal de sodio voltaje operado) o los receptores de neurotransmisores (como el receptor de ácido- $\gamma$ -aminobutírico de tipo A -GABAA).



**Figura 2.3.** Representación esquemática de la hipótesis del blanco molecular. **A.** Blanco molecular normal al cual se une el FAE para desencadenar la posterior actividad farmacológica deseada y, **B.** Blanco molecular alterado en un paciente epiléptico refractario al tratamiento, donde el FAE no puede unirse a su blanco molecular y por lo tanto no se observará el efecto terapéutico esperado.

La hipótesis de la variante genética sugiere que hay una resistencia inherente que se rige por las variantes genéticas de las proteínas que están implicadas en la farmacocinética y la farmacodinamia de los FAE, como por ejemplo variantes polimórficas de las enzimas de biotransformación de los fármacos. Algunos años atrás, las alteraciones constitutivas de los transportadores y de los blancos moleculares de los FAE se consideraban dentro del alcance de la hipótesis de los transportadores y del blanco molecular, respectivamente [19, 24]. Recientemente, expertos de referencia en el campo parecen preferir categorizar las alteraciones intrínsecas (variantes genéticas) del blanco molecular dentro de la hipótesis de la variante genética [10]. En el marco de este nuevo esquema de clasificación, sólo modificaciones adquiridas en los transportadores o blancos moleculares se encontrarán dentro de la hipótesis de los transportadores o del blanco molecular. Aunque a primera vista esta distinción puede parecer como un problema de clasificación trivial, la naturaleza de la alteración farmacocinética o farmacodinámica podría tener un profundo impacto en el abordaje clínico del problema de refractariedad en la epilepsia. Mientras que una causa genética de farmacoresistencia actualmente podría ser detectada a través de pruebas de diagnóstico simples, incluso antes de iniciar el tratamiento, modificaciones adquiridas vinculadas a la fisiopatología del trastorno son más difíciles de probar y

hasta podrían requerir procedimientos invasivos como por ejemplo, de cirugía de resección.

La hipótesis de las redes neuronales sostiene que los episodios recurrentes de excesiva actividad neuronal conducen a alteraciones plásticas y estructurales del cerebro y a una reorganización de la red neuronal. La generación de redes anormales podrían a su vez estar vinculadas con el fenómeno MDR en la epilepsia. Esta hipótesis es apoyada por el hecho de que la resección quirúrgica del foco epiléptico conduce con frecuencia a la ausencia de crisis convulsivas o revierte el estado refractario a los FAE [10]. De hecho, las epilepsias de causa estructural están vinculadas a farmacorresistencia y a imágenes cerebrales anormales, y un buen resultado quirúrgico (es decir, ausencia de crisis) está asociado a una resección quirúrgica completa del foco o lesión epiléptica [25]. Las diferencias entre las alteraciones en la plasticidad del cerebro en pacientes responsivos y no responsivos aún no se han establecido [20].

Por último, la hipótesis de la severidad intrínseca se basa en estudios epidemiológicos que muestran que la gravedad inicial del cuadro, es decir el número de convulsiones en las primeras manifestaciones de este trastorno, condiciona su pronóstico y conduce a una mayor dificultad para controlar las crisis [10]. De nuevo, las bases biológicas de la severidad intrínseca no son completamente entendidas hasta el momento [26], por lo que la influencia de esta hipótesis en la elección y/o desarrollo del tratamiento es actualmente limitada.

Recientemente, se ha sugerido un posible papel de la epigenética en la epilepsia refractaria, estableciendo una sexta hipótesis para la epilepsia refractaria que podría servir para ampliar la base biológica de algunas de las anteriores (tales como la hipótesis de los transportadores y la del blanco molecular), aunque hasta el momento la evidencia experimental que apoya esta hipótesis sigue siendo escasa [27].

Estas hipótesis no son mutuamente excluyentes y ninguna proporciona una explicación completa o universal para los pacientes con epilepsia refractaria: una hipótesis dada podría ser aplicable a un subgrupo particular de pacientes o, alternativamente, algunos pacientes podrían presentar varios mecanismos al mismo tiempo y por lo tanto requerir múltiples hipótesis para explicar su falta de respuesta al tratamiento [10, 24, 28]. La hipótesis de las redes neuronales parece,

hasta el momento, la explicación más global de la MDR en la epilepsia, ya que algunas de las otras explicaciones (por ejemplo, la hipótesis del blanco molecular) podrían ser aplicadas dentro de su contexto [10]. Merece la pena subrayar que el enfoque del tratamiento debe ser altamente dependiente de los mecanismos de MDR presentes en cada paciente en particular. Superar la refractariedad representa un gran reto y requerirá de un enorme esfuerzo conjunto de investigadores de todos los niveles, desde las ciencias básicas hasta la clínica, siendo la retroalimentación de la información el pilar fundamental para lograrlo.

### ***2.3 Hipótesis de los transportadores: la evidencia***

Como se mencionó en la sección anterior, una de las hipótesis más estudiadas y aceptadas para explicar la MDR en la epilepsia es la hipótesis de los transportadores, y esto se debe a la gran cantidad de evidencia experimental y clínica acumulada a lo largo de los años que soporta dicha hipótesis. La evidencia demuestra la presencia de altos niveles de expresión de los transportadores MDR-ABC en la unidad neurovascular tanto en los modelos animales de epilepsia refractaria como en los pacientes refractarios, sobreexpresión que ha sido asociada a la activación transcripcional ligada tanto a la propia fisiopatología como a la farmacoterapia. En relación a esta última, se ha descrito que la expresión de varios miembros de la superfamilia ABC es inducida a través del receptor X de pregnano [29-31], el cual es activado por una gran diversidad de compuestos endógenos y exógenos [30]. En este sentido, se demostró que ciertos FAE son capaces de inducir la expresión de los transportadores MDR-ABC [32-33]. En relación a la fisiopatología, la actividad convulsiva demostró actuar como un fuerte disparador del aumento de los niveles de expresión de Pgp y otros transportadores MDR-ABC a nivel de la unidad neurovascular [18, 34-41].

Sin embargo, no toda la evidencia acumulada hasta el momento es favorable, sino que existe cierta cantidad de evidencia aparentemente conflictiva particularmente en términos clínicos (Tabla 2.1) que se analizará en detalle a continuación.

**Tabla 2.1.** Resumen de la evidencia acumulada en términos preclínicos y clínicos [23].

Hipótesis de los transportadores	Evidencia preclínica	Evidencia clínica
Los transportadores están presentes en el tejido cerebral epileptogénico	+	+
Los transportadores en el tejido epileptogénico presentan la funcionalidad adecuada para mediar la resistencia a los FAE	+	?
Los transportadores se encuentran implicados en la resistencia a los FAE <i>in vivo</i>	+	?
Inhibir la acción de los transportadores contrarresta la resistencia a los FAE	(+)	(+)

+ ha sido demostrado; ? aún no se ha podido demostrar; (+) hay evidencia conflictiva

### 2.3.1 Evidencia preclínica

La hipótesis de los transportadores ha sido corroborada a nivel preclínico, ya que la resistencia a los fármacos en modelos animales de epilepsia refractaria se ha revertido por la administración conjunta de inhibidores de los transportadores MDR-ABC. En el año 2001, Potschka y col. [42] dieron los primeros pasos hacia una prueba de concepto de la posible función de los transportadores MDR-ABC en este fenómeno: mediante el uso de microdiálisis en ratas *in vivo*, los autores mostraron que los niveles de carbamazepina en el fluido extracelular de la corteza cerebral podían aumentarse a través de la perfusión local del inhibidor de Pgp verapamilo y del inhibidor de MRP1, MRP2, MRP5 y MRP6 probenecid. Algún tiempo después, los mismos investigadores demostraron que la administración conjunta de fenitoína a dosis sub-anticonvulsivantes y probenecid resultó en un claro aumento del efecto anticonvulsivo de fenitoína en ratas encendidas eléctricamente, asociado a un aumento significativo de los niveles del fármaco en el cerebro [43]. Más tarde

se obtuvieron resultados similares en el modelo focal de pilocarpina de convulsiones límbicas en ratas [44].

Si bien estos trabajos pioneros fueron muy valiosos, tenían dos limitaciones importantes: por un lado, el uso de moduladores de los transportadores ABC de primera generación (débiles e inespecíficos), y por otro lado que los experimentos fueron realizados sin discriminar entre aquellos animales que fueran respondedores al tratamiento y aquellos no respondedores o refractarios.

La introducción de tariquidar, un inhibidor de tercera generación específico de Pgp, resolvió la primera cuestión. Utilizando un modelo de rata de epilepsia crónica van Vliet y col. [45] demostraron que, mientras que la administración de fenitoína sola no conseguía la supresión completa de las convulsiones espontáneas, la co-administración de fenitoína y tariquidar condujo a un control de las convulsiones casi completo, asociado al aumento de la relación cerebro-a-plasma de fenitoína consecuencia de la inhibición de la Pgp mediada por tariquidar. Sin embargo, el efecto de la co-administración de tariquidar en el control de las convulsiones se revirtió después de cuatro días, lo que sugiere el desarrollo de tolerancia a tariquidar. La prueba de concepto preclínica definitiva se obtuvo mediante la co-administración de tariquidar a ratas con epilepsia del lóbulo temporal farmacorresistente con probada asociación a la regulación hacia arriba de la expresión de Pgp [46]. Se utilizó un protocolo para discriminar entre animales respondedores y no respondedores a fenobarbital y cuando se co-administró tariquidar, cinco de cada seis animales no respondedores resultaron libres de crisis o mostraron una reducción de las crisis de al menos el 50%. Se obtuvieron resultados similares en el modelo de epilepsia refractaria de ácido 3-mercaptopropiónico, que se asocia a regulación hacia arriba de Pgp en la BHE, los astrocitos y las neuronas [47]. El pre-tratamiento de las ratas epilépticas por el modelo de ácido 3-mercaptopropiónico con el inhibidor de Pgp nimodipina condujo a una mayor biodisponibilidad de fenitoína a nivel del hipocampo. Posteriormente, en un estudio sobre 11 perros resistentes a fenobarbital el tratamiento adicional con verapamilo no logró un aumento en el control de las convulsiones, e incluso algunos animales mostraron una tendencia a empeorar el control de las crisis [48]. Estos resultados destacan la importancia de las diferencias entre especies y la necesidad de validar la hipótesis de los

transportadores a nivel clínico. Aun así, la prueba de concepto clínica sigue siendo difícil de alcanzar, como se analizará a continuación.

### *2.3.2 Evidencia clínica*

En cuanto a los datos clínicos, se ha acumulado una gran cantidad de evidencia que demuestra altos niveles de expresión de los transportadores ABC (particularmente de Pgp, BCRP y varios miembros de las MRP) en la unidad neurovascular, ya sea a nivel de la BHE, las células de la glía o las neuronas de pacientes refractarios al tratamiento [17, 23, 34-36, 38-39, 49-55]. Sin embargo, en la mayoría de estos estudios se compararon muestras de cerebro de pacientes con epilepsia refractaria que habían sido sometidos a la extirpación quirúrgica del foco epiléptico con muestras de cerebro humano sin antecedentes de convulsiones. Si bien el tejido cerebral de pacientes epilépticos respondedores sería un control más adecuado, tales muestras control por lo general no están disponibles ya que el procedimiento invasivo que se requiere para conseguir las es éticamente inaceptable en pacientes respondedores. Esta limitación se ha superado en estudios más recientes donde usando tomografía por emisión de positrones (PET) [56-57], se demostró que la constante de velocidad de transporte plasma a cerebro para [<sup>11</sup>C]verapamil tiende a ser menor en diferentes regiones del cerebro de pacientes epilépticos refractarios a los FAE en comparación con pacientes libres de crisis e individuos sanos, y que luego de la administración de valproato dicha constante se incrementa en todo el cerebro tanto de sujetos sanos como de pacientes refractarios. Este estudio mostró que cerca del 40% de los pacientes refractarios presentaban un aumento de la funcionalidad de la Pgp en el foco epiléptico, por lo que son pacientes susceptibles de beneficiarse de la inhibición de este transportador.

Algunos estudios han demostrado concentraciones reducidas de los FAE en el fluido extracelular cerebral y en el tejido epileptogénico de pacientes no responsivos [58-59]. Si bien estos son pasos muy importantes hacia la validación de la hipótesis de los transportadores, una prueba definitiva requeriría la inversión de la resistencia a los fármacos (es decir, lograr el control de las convulsiones)

después de bloquear los transportadores MDR-ABC. Se han reportado algunos casos anecdóticos de pacientes refractarios que han mostrado mejoría cuando los FAE fueron co-administrados con verapamilo [60-63], pero aún no está claro si los resultados observados podrían ser una consecuencia de la actividad anticonvulsiva intrínseca de verapamilo (por ejemplo, a través de la modulación de la entrada de calcio en las neuronas) u otro efecto de este fármaco sobre la farmacocinética de los FAE. Recientemente se llevó a cabo un estudio piloto en siete niños con diferentes diagnósticos de epilepsia refractaria los cuales recibieron verapamilo como tratamiento adicional al FAE de base. Los resultados fueron variados desde ausencia de respuesta a respuesta parcial, la cual en algunos pacientes fue seguida de un empeoramiento de las convulsiones [64]. Aunque el número de pacientes que participaron en el estudio fue muy limitado, los resultados están en línea con la idea de que algunas intervenciones terapéuticas podrían ser más eficaces en determinados subgrupos de pacientes no respondedores [65].

Sigue siendo necesaria la realización de ensayos clínicos multicéntricos aleatorizados y estudios relacionados al efecto de los inhibidores selectivos de la Pgp sin actividad intrínseca para obtener evidencia clínica definitiva de la hipótesis de los transportadores. En cuanto a la asociación entre variantes genéticas de los transportadores ABC y la epilepsia resistente a los FAE, los resultados de estudios son controvertidos o no concluyentes; mientras que meta-análisis previos no lograron establecer una asociación entre las variantes de ABCB1 y la epilepsia refractaria [66], publicaciones más recientes han sugerido asociaciones en sujetos asiáticos y caucásicos [67-69] contribuyendo a la validez de la hipótesis.

### *2.3.3 Transportadores MDR-ABC: un trabajo en equipo*

Si bien varios FAE importantes, como fenitoína, fenobarbital, topiramato, levetiracetam, oxcarbazepina y lamotrigina son sustratos de la Pgp humana [12, 70], el principal argumento en contra de la hipótesis de los transportadores parece ser que no todos los FAE lo son. Aunque existe evidencia aparentemente conflictiva respecto a cuáles FAE son sustratos y cuáles no [12, 70], los resultados dependen en gran medida del sistema experimental, en especial del tipo de ensayo (*in vivo*, *ex*

*vivo* o *in vitro*; modelos de células humanas *versus* modelos de células animales; ensayo de transporte a la concentración de equilibrio o en condiciones de no equilibrio) y de la posible variabilidad entre especies en la especificidad de sustrato; por lo que se deben tener en cuenta algunas consideraciones antes de arribar a una conclusión definitiva sobre la categoría de un fármaco.

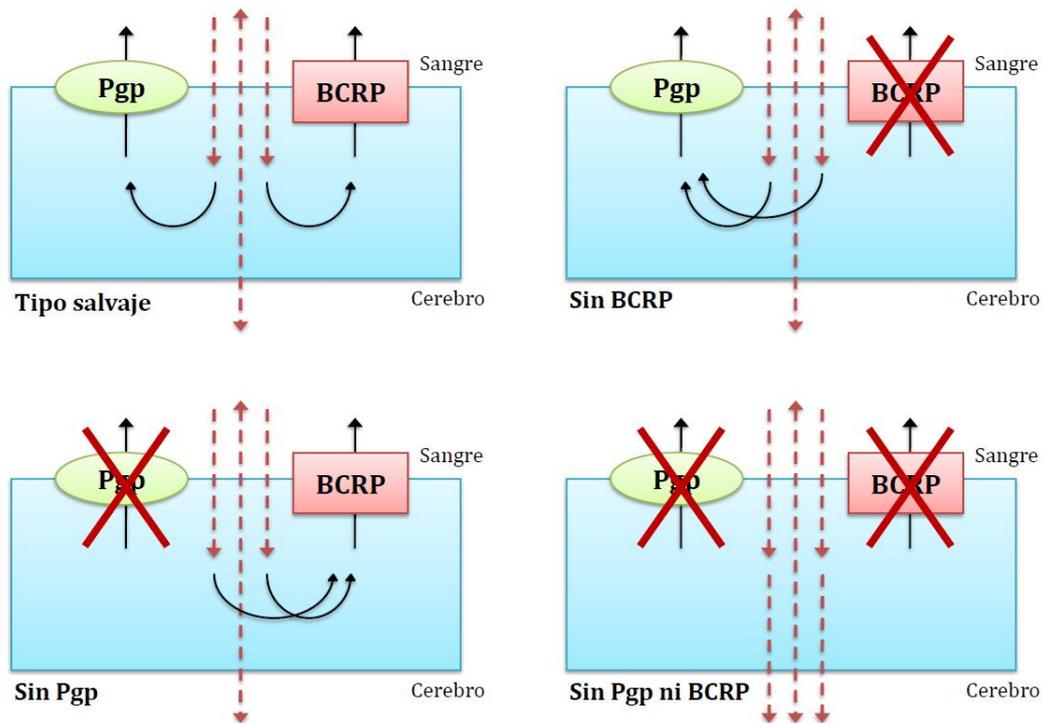
En relación con los ensayos de permeabilidad *in vitro*, los ensayos de transporte bi-direccional en presencia y ausencia de un inhibidor selectivo de Pgp pueden carecer de sensibilidad en algunos casos, ya que el transporte direccional por el transportador podría estar enmascarado por la contribución de la difusión pasiva. La magnitud de este efecto depende de las concentraciones ensayadas del fármaco, de los niveles de expresión del transportador en el cultivo celular, de la afinidad entre el fármaco y el transportador y de las características fisicoquímicas del fármaco (por ejemplo, la permeabilidad), entre otros factores [71]. Comenzar el ensayo con concentraciones idénticas del fármaco en ambos lados de la monocapa de células (ensayo de transporte a la concentración de equilibrio, CETA) mejora la sensibilidad del ensayo ya que elimina el gradiente de concentración, eliminando la difusión neta y permitiendo evaluar el transporte activo independientemente de la componente de la permeabilidad pasiva que puede causar que los fármacos altamente lipofílicos (como la mayoría de los FAE) no sean identificados como sustratos del transportador [12, 72].

Por otro lado, el hecho de que se haya demostrado en modelos apropiados que algunos FAE no son sustratos de Pgp no elimina en absoluto a la hipótesis de los transportadores como potencial explicación de la resistencia a fármacos. Si bien la mayor parte de la investigación sobre los transportadores MDR-ABC en epilepsia refractaria (y en general en todos los casos de MDR) se ha centrado en la Pgp (el primer miembro identificado de la superfamilia) algunos FAE son reconocidos y transportados por otros transportadores MDR-ABC.

Entre los transportadores MDR-ABC regulados hacia arriba en pacientes refractarios se han reportado varios miembros de las MRP, en particular MRP1, MRP2, MRP5 y MRP6 [37-39], aunque según los resultados obtenidos mediante ensayos *in vitro* no transportan significativamente ninguno de los FAE más comúnmente utilizados [73-74]. Más recientemente, se ha reportado la sobreexpresión de la BCRP humana en la epilepsia refractaria y su papel en los

fenómenos de resistencia a los FAE podría haber sido subestimado: si bien los primeros trabajos parecían sugerir que varios FAE no eran reconocidos ni transportados por BCRP *in vitro* [72], en estudios *in vivo* mediante el uso de ratones modificados genéticamente que carecen ya sea de Pgp o de Pgp y Bcrp, se evidenció que ambos transportadores restringen el acceso al cerebro de varios FAE [75], mientras que los últimos estudios publicados hasta el momento demostraron mediante el ensayo CETA que lamotrigina es de hecho un sustrato tanto de Pgp como de BCRP [74]. Cabe destacar aquí que mediante estudios de proteómica se ha observado que la BCRP es el transportador con los niveles de expresión basales más altos tanto a nivel de la BHE [76-77] como en todos los segmentos del intestino humano de sujetos sanos [78-79], además de ser el transportador ABC con espectro de sustratos más amplio [80].

Analizar individualmente el papel de los transportadores ABC en los fenómenos de MDR puede no ser el camino correcto dado que la evidencia acumulada sugiere una superposición parcial de la especificidad de sustrato entre diferentes miembros de la superfamilia sumada a patrones de co-expresión y co-localización, donde el aumento o disminución de la expresión o actividad de un transportador determinado puede verse equilibrado por una regulación compensatoria en los niveles de expresión de otros transportadores de la superfamilia. Todo apunta a una función de cooperación en la disposición de sustratos comunes a los distintos transportadores [81-85]. Hay evidencia que señala que la cooperación entre Pgp y BCRP implica que la sola ausencia de una de las dos proteínas no da lugar a un aumento significativo de la distribución al cerebro de los sustratos duales: en ratones que carecen de la Pgp, la sola presencia de la BCRP es suficiente para reducir la biodisponibilidad del fármaco en el cerebro, y viceversa. El mayor aumento de la biodisponibilidad en el cerebro de los sustratos duales siempre se obtiene cuando tanto Pgp como BCRP están ausentes (Figura 2.4) [86]. Esto implica que el papel de un cierto transportador ABC podría ser ocultado por la función de los demás, lo que subraya la necesidad de evaluar el posible reconocimiento de los FAE por parte de otros transportadores ABC aparte de Pgp. Además, se ha visto que el efecto de los transportadores ABC suele ser sinérgico (y no aditivo) en los casos donde el fármaco es reconocido y transportado por más de un transportador a la vez [87-88].



**Figura 2.4.** Representación de la cooperación entre Pgp y BCRP a nivel de la BHE para restringir la penetración en el cerebro de los sustratos duales. La ausencia de cualquiera de los dos transportadores por sí sola puede no dar lugar a un aumento significativo en la penetración al cerebro de los sustratos duales; es decir que en ratones que carecen de BCRP (donde Pgp está presente) la regulación hacia arriba de Pgp podría ser suficiente para evitar la entrada de sustratos duales al cerebro, y viceversa. El aumento en la penetración al cerebro de los sustratos duales se ve más claramente en los ratones que carecen de ambos transportadores. *Adaptado de [86].*

A pesar de que la investigación experimental y clínica abordó mediante múltiples estudios el rol de los transportadores MDR-ABC en el contexto de la epilepsia refractaria, aún no se ha logrado la prueba de concepto definitiva en cuanto a su relevancia clínica [89]. Las dificultades para cuantificar los niveles de expresión de un transportador dado en diferentes regiones del cerebro de un paciente epiléptico que no ha sido sometido a una intervención quirúrgica, y la incertidumbre con respecto a la capacidad de los modelos experimentales para reproducir los niveles de expresión absoluta y relativa de los diferentes transportadores de eflujo MDR-ABC en el foco epiléptico y la BHE (niveles de expresión que bien podrían ser altamente dinámicos y dependientes del paciente), contribuyen a la dificultad de evaluar de manera inequívoca la influencia de un

determinado transportador en la biodisponibilidad de los FAE a nivel cerebral. Es necesario el desarrollo de nuevos estudios clínicos para ampliar la evaluación de la correlación entre las contribuciones separadas y concertadas de los diferentes transportadores MDR-ABC y la respuesta clínica a los FAE.

Finalmente, cabe destacar que la definición actual de epilepsia refractaria no obliga a ensayar fármacos no sustrato de transportadores MDR-ABC como alternativa de tratamiento; la definición solamente contempla la persistencia de las crisis luego de dos intervenciones farmacológicas adecuadas y bien toleradas, aunque la calidad de “adecuada” nada dice de la afinidad de los FAE ensayados como tratamiento por los transportadores MDR-ABC. Por lo tanto, es dado que las dos intervenciones terapéuticas necesarias para diagnosticar epilepsia refractaria sean fármacos sustratos de la transportadores MDR-ABC, sin invalidar la hipótesis el hecho de que existan FAE de uso clínico que no sean sustratos de los mismos.

#### ***2.4 Estrategias para superar la MDR mediada por transportadores ABC en la epilepsia refractaria***

A partir de la fuerte aceptación del rol de los transportadores en fenómenos de MDR, comenzaron a explorarse una serie de posibles soluciones terapéuticas. Los mayores esfuerzos se centraron en la búsqueda de inhibidores de los transportadores MDR-ABC (Tabla 2.2) [90-102]. Los primeros trabajos en este sentido se iniciaron con la co-administración de inhibidores como una posible solución terapéutica para la farmacorresistencia en el cáncer, donde luego de algunos éxitos iniciales como en el tratamiento de retinoblastoma mediante la combinación de agentes quimioterapéuticos con ciclosporina para inhibir la Pgp [103], los estudios clínicos llevados a cabo en el área de los anticancerígenos con inhibidores de P-gp de primera (no selectivos y con actividad farmacológica intrínseca), segunda (mayor selectividad y menor actividad farmacológica intrínseca) e incluso tercera generación (altamente selectivos y con menor o ninguna actividad farmacológica intrínseca) han sido decepcionantes dado que indican que la inhibición de estos transportadores puede conducir a reacciones adversas severas [89-90, 104-106]. Pese a que la búsqueda de nuevos y mejores

inhibidores (particularmente de inhibidores más selectivos) continúa [107-114], debe tenerse en cuenta el papel fisiológico de los transportadores ABC como componentes fundamentales de los sistemas de detoxificación general y protección del organismo, además de su participación en el tráfico de compuestos endógenos. Esto hace que su inhibición como complemento al tratamiento farmacológico no sea segura, en particular en aquellos casos en los que se requiere un tratamiento a largo plazo, como es el caso de la epilepsia y el cáncer. La conexión entre la disfunción de los transportadores ABC y enfermedades neurodegenerativas como el Parkinson y la enfermedad de Alzheimer debe ser tenida en cuenta como un ejemplo del riesgo potencial que representa la inhibición crónica de estos sistemas de eflujo [115-117]. Se ha probado además que la Pgp regula el transporte de glutamato a nivel de la BHE en condiciones normales [118], por lo que podría estar implicada en la regulación de otros neurotransmisores. Los efectos potenciales de estos inhibidores en la farmacocinética de otros medicamentos también deben ser considerados en un escenario de pacientes polimedicados, debido a la alta probabilidad de interacciones farmacológicas adversas. La superposición de la especificidad de sustrato y las posibilidades de efectos compensatorios (donde la pérdida o inhibición de un determinado transportador ABC dispara la regulación hacia arriba de un transportador alternativo), suponen un desafío aún mayor en el camino para lograr un beneficio terapéutico. Inhibidores débiles de los transportadores MDR-ABC emergen como posibles soluciones, así como también agentes terapéuticos dirigidos a blancos moleculares vinculados a la cascada de señalización que regula la expresión de los transportadores MDR-ABC en el cerebro epiléptico [33, 89]. Estas dos últimas opciones pueden ser útiles para prevenir o moderar la regulación hacia arriba de la función de los transportadores MDR-ABC inducida por drogas o por el propio trastorno mediante la activación de los receptores nucleares o a través de señales pro-inflamatorias, respectivamente [89, 119].

En definitiva, la inhibición de estos transportadores como parte de un tratamiento crónico determina un escenario clínico complejo y sigue siendo tema de investigación y debate.

**Tabla 2.2.** Listado de los inhibidores más relevantes de los transportadores MDR-ABC involucrados en epilepsia refractaria [23, 120].

Transportador MDR-ABC	Inhibidores/moduladores	
<b>Pgp</b>	<b>Primera generación</b>	Verapamilo, nifedipina, ciclosporina A, quinidina, quinina, amiodarona, trifluoperazina, trifluopromazina, clorpromazina, prometazina, azidopina, flufenazina, pimozina, ketoconazol, tamoxifeno, toremifeno, reserpina, cefalosporinas, yohimbina, propranolol, detergentes como el cremophor EL, Tween 80, Triton X-100
	<b>Segunda generación</b>	PSC-833 (valspodar), GF120918 (elacridar), VX-710 (biricodar), dexverapamilo, dexniguldipina, dofequidar fumarato (MS-209)
	<b>Tercera generación</b>	OC-144-093 (ONT-093), LY335979 (zosuquidar), XR9576 (tariquidar), R101933 (laniquidar), GF120918 (elacridar), S9788, NSC-38721 (mitotano), HM30181, WK-X-34
<b>BCRP</b>	GF120918 (elacridar), VX-710 (biricodar), imatinib mesilato, ciclosporina A, TAG-139, WK-X-34, fumitremorgina C (FTC) y sus análogos como el Ko132, Ko134 y Ko143, CI-1033, pantoprazol	
<b>MRP1</b>	Sulfinpirazona, probenecid, MK-571, LTC4, MS-209, benzbromarona, SCH66336 (lonafarnib), agosterol A, rifampicina, isotiocianatos, ciclosporina A, verapamilo, quinidina, quinina, PSC 833, biricodar (VX-710)	
<b>MRP2</b>	Leucotrieno C4, probenecid, MK-571, biricodar, isotiocianatos	
<b>MRP5</b>	Probenecid e inhibidores de fosfodiesterasas como trequinsin o sildenafil	

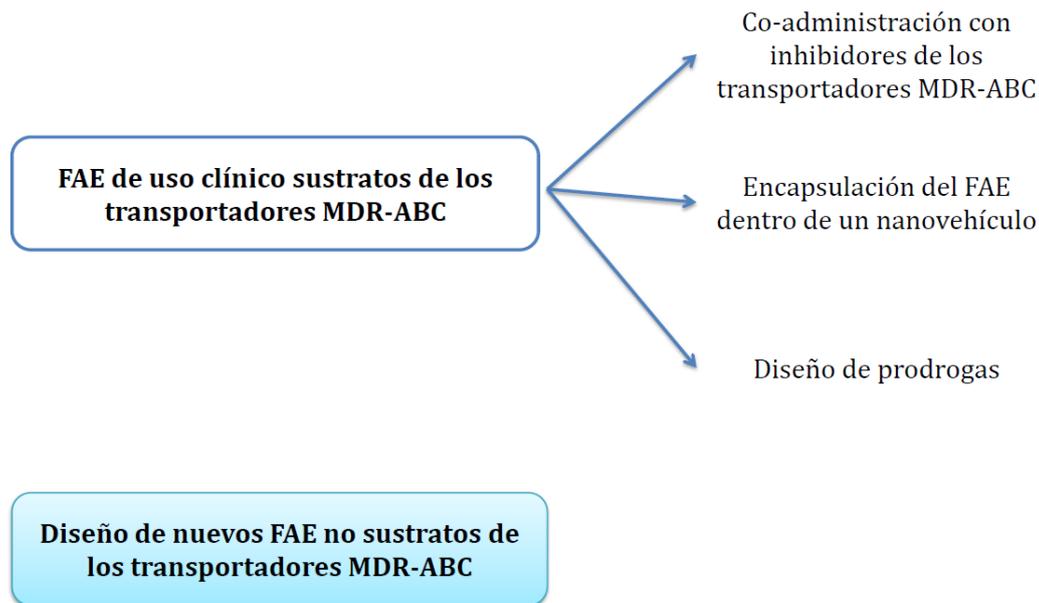
<b>MRP6</b>	Probenecid, indometacina e inhibidores de fosfodiesterasas como trequinsin o sildenafil
-------------	---

Por otro lado puede mencionarse el uso de una estrategia de tipo “caballo de Troya” para lograr alcanzar niveles terapéuticos de los sustratos de los transportadores MDR-ABC en el foco epiléptico, evitando el reconocimiento de los mismos por parte de los transportadores. Mientras que la administración local de los FAE suele descartarse por el carácter invasivo de la técnica [121], acarreadores farmacéuticos particulados como los nanovehículos y otros vectores de última generación resultan prometedores y están siendo arduamente estudiados [122-124]. Este enfoque conlleva la encapsulación del FAE de uso clínico sustrato de los transportadores MDR-ABC en sistemas de liberación avanzados como las nanopartículas. Siempre que se utilicen vectores seguros, esta estrategia parecería presentar mejores posibilidades de sobrevivir a los ensayos clínicos.

En consonancia con el enfoque anterior también podemos mencionar el diseño de profármacos de los FAE, es decir, enmascarar químicamente fármacos sustrato de los transportadores MDR-ABC para que carezcan de afinidad y no sean reconocidos por estos transportadores, o para que muestren afinidad por transportadores de afluencia que podrían compensar el efecto de las bombas de eflujo en la permeabilidad a través de la BHE. Aunque se han desarrollado varios profármacos de FAE aprobados, la interacción de la mayoría de ellos con los transportadores de eflujo aún no fue evaluada [119, 123].

Finalmente otra solución alternativa muy interesante pero aún poco explorada consiste en la detección temprana de sustratos de los transportadores MDR-ABC durante la primeras etapas de desarrollo de nuevos fármacos con la finalidad de descartar aquellos que son potenciales sustratos y avanzar en las etapas futuras de investigación con aquellos que tengan mayores chances de no ser reconocidos por estos transportadores (es decir, capaces de eludirlos) y por lo tanto no presenten problemas de MDR mediada por los mismos.

Un esquema que ilustra las diferentes estrategias se presenta en la Figura 2.5.



**Figura 2.5.** Resumen de las estrategias terapéuticas propuestas hasta el momento para superar la epilepsia refractaria mediada por los transportadores MDR-ABC

### *2.4.1 Aproximación aplicada en la tesis*

Del análisis anterior se desprende que el desarrollo de nuevos FAE que no sean reconocidos por los transportadores MDR-ABC y el encapsulamiento de fármacos sustratos en nanovehículos parecen ser las dos soluciones terapéuticas más aceptables y con mayor probabilidad de éxito para abordar el fenómeno de epilepsia refractaria mediada por los transportadores MDR-ABC. Mientras que la segunda opción está siendo intensamente estudiada, la primera resulta más novedosa y menos explorada.

Si bien la Pgp es el transportador de eflujo más ampliamente investigado, los recientes avances expuestos aquí sugieren que el papel de la BCRP en el fenómeno de la epilepsia refractaria no debe subestimarse, por lo que el reconocimiento temprano de sustratos de la Pgp debe ser necesariamente complementado con el reconocimiento temprano de sustratos de la BCRP durante las primeras etapas del diseño de fármacos, con el fin de evitar el reconocimiento mediado por estos dos transportadores y así poder obtener niveles terapéuticos de los FAE en el cerebro

de los pacientes refractarios. Por otra parte, la Administración de Medicamentos y Alimentos de los Estados Unidos (FDA) ha reconocido a la BCRP como uno de los transportadores de fármacos de mayor importancia clínica y recomienda que todos los fármacos en investigación deben ser evaluados *in vitro* para determinar si son un potencial sustrato y/o inhibidor tanto de la Pgp como de la BCRP [110, 125-126].

Por todo esto, en la presente tesis se plantea el desarrollo de modelos computacionales (los cuales serán analizados en detalle en el siguiente capítulo) que permitan el reconocimiento temprano de sustratos de la BCRP para asistir en la búsqueda de nuevos FAE que no sean reconocidos por este transportador y que por lo tanto no presenten problemas de farmacorresistencia mediada por BCRP, aumentando así la probabilidad de éxito terapéutico. Cabe destacar que los modelos aquí desarrollados no sólo presentan utilidad para el diseño de nuevos FAE para la epilepsia refractaria, sino que pueden ser aplicados también en etapas tempranas del desarrollo de nuevos fármacos para el tratamiento de otras patologías con problemas de MDR mediada por BCRP y para la predicción de potenciales interacciones medicamentosas entre dos o más fármacos sustratos de dicho transportador.

---

## ***Referencias Capítulo II***

- [1] Thurman DJ, Beghi E, Begley CE, Berg AT, Buchhalter JR, Ding D, et al. Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia*. 2011;52 Suppl 7:2-26.
- [2] World Health Organization. Fact sheet number 999: Epilepsy. 2015 [updated May, 2015; Accessed August 19, 2015]; Available from: <http://www.who.int/mediacentre/factsheets/fs999/en/>
- [3] Carpio A, Hauser WA. Epilepsy in the developing world. *Curr Neurol Neurosci Rep*. 2009;9(4):319-26.
- [4] Moshe SL, Perucca E, Ryvlin P, Tomson T. Epilepsy: new advances. *Lancet*. 2015;385(9971):884-98.
- [5] Brodie MJ, Barry SJ, Bamagous GA, Norrie JD, Kwan P. Patterns of treatment response in newly diagnosed epilepsy. *Neurology*. 2012;78(20):1548-54.
- [6] Kwan P, Arzimanoglou A, Berg AT, Brodie MJ, Allen Hauser W, Mathern G, et al. Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies. *Epilepsia*. 2010;51(6):1069-77.
- [7] Sinha S, Siddiqui KA. Definition of intractable epilepsy. *Neurosciences (Riyadh)*. 2011;16(1):3-9.
- [8] Schiller Y, Najjar Y. Quantifying the response to antiepileptic drugs: effect of past treatment history. *Neurology*. 2008;70(1):54-65.
- [9] Loscher W, Schmidt D. Modern antiepileptic drug development has failed to deliver: ways out of the current dilemma. *Epilepsia*. 2011;52(4):657-78.
- [10] Loscher W, Klitgaard H, Twyman RE, Schmidt D. New avenues for anti-epileptic drug discovery and development. *Nat Rev Drug Discov*. 2013;12(10):757-76.
- [11] Kwan P, Brodie MJ. Refractory epilepsy: mechanisms and solutions. *Expert Rev Neurother*. 2006;6(3):397-406.
- [12] Loscher W, Luna-Tortos C, Romermann K, Fedrowitz M. Do ATP-binding cassette transporters cause pharmaco-resistance in epilepsy? Problems and approaches in determining which antiepileptic drugs are affected. *Curr Pharm Des*. 2011;17(26):2808-28.

- [13] Schmidt D, Schachter SC. Drug treatment of epilepsy in adults. *BMJ*. 2014;348:g254.
- [14] Shorvon SD. Drug treatment of epilepsy in the century of the ILAE: the first 50 years, 1909-1958. *Epilepsia*. 2009;50 Suppl 3:69-92.
- [15] Shorvon SD. Drug treatment of epilepsy in the century of the ILAE: the second 50 years, 1959-2009. *Epilepsia*. 2009;50 Suppl 3:93-130.
- [16] Löscher W. Mechanisms of drug resistance. *Epileptic Disorders*. 2005;7(Suppl 1):S3-S9.
- [17] Kwan P, Brodie MJ. Potential role of drug transporters in the pathogenesis of medically intractable epilepsy. *Epilepsia*. 2005;46(2):224-35.
- [18] Remy S, Beck H. Molecular and cellular mechanisms of pharmacoresistance in epilepsy. *Brain*. 2006;129(Pt 1):18-35.
- [19] Schmidt D, Löscher W. Drug resistance in epilepsy: putative neurobiologic and clinical mechanisms. *Epilepsia*. 2005;46(6):858-77.
- [20] Fang M, Xi ZQ, Wu Y, Wang XF. A new hypothesis of drug refractory epilepsy: neural network hypothesis. *Med Hypotheses*. 2011;76(6):871-6.
- [21] Rogawski MA, Johnson MR. Intrinsic severity as a determinant of antiepileptic drug refractoriness. *Epilepsy Curr*. 2008;8(5):127-30.
- [22] Löscher W, Klotz U, Zimprich F, Schmidt D. The clinical impact of pharmacogenetics on the treatment of epilepsy. *Epilepsia*. 2009;50(1):1-23.
- [23] Löscher W, Potschka H. Role of drug efflux transporters in the brain for drug disposition and treatment of brain diseases. *Prog Neurobiol*. 2005;76(1):22-76.
- [24] Schmidt D, Löscher W. New developments in antiepileptic drug resistance: an integrative view. *Epilepsy Curr*. 2009;9(2):47-52.
- [25] Wiebe S, Jette N. Pharmacoresistance and the role of surgery in difficult to treat epilepsy. *Nat Rev Neurol*. 2012;8(12):669-77.
- [26] Rogawski MA. The intrinsic severity hypothesis of pharmacoresistance to antiepileptic drugs. *Epilepsia*. 2013;54 Suppl 2:33-40.
- [27] Kobow K, El-Osta A, Blumcke I. The methylation hypothesis of pharmacoresistance in epilepsy. *Epilepsia*. 2013;54 Suppl 2:41-7.
- [28] Talevi A, Bruno-Blanch LE. On the Development of New Antiepileptic Drugs for the Treatment of Pharmacoresistant Epilepsy: Different Approaches to Different Hypothesis. In: Rocha L, Cavalheiro EA, editors. *Pharmacoresistance in*

Epilepsy From Genes and Molecules to Promising Therapies. New York: Springer-Verlag; 2013. p. 207-24.

[29] Geick A, Eichelbaum M, Burk O. Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *J Biol Chem.* 2001;276(18):14581-7.

[30] Miller DS, Bauer B, Hartz AM. Modulation of P-glycoprotein at the blood-brain barrier: opportunities to improve central nervous system pharmacotherapy. *Pharmacol Rev.* 2008;60(2):196-209.

[31] Bauer B, Hartz AM, Lucking JR, Yang X, Pollack GM, Miller DS. Coordinated nuclear receptor regulation of the efflux transporter, Mrp2, and the phase-II metabolizing enzyme, GSTpi, at the blood-brain barrier. *J Cereb Blood Flow Metab.* 2008;28(6):1222-34.

[32] Wen T, Liu YC, Yang HW, Liu HY, Liu XD, Wang GJ, et al. Effect of 21-day exposure of phenobarbital, carbamazepine and phenytoin on P-glycoprotein expression and activity in the rat brain. *J Neurol Sci.* 2008;270(1-2):99-106.

[33] Potschka H. Targeting regulation of ABC efflux transporters in brain diseases: a novel therapeutic approach. *Pharmacol Ther.* 2010;125(1):118-27.

[34] Tishler DM, Weinberg KI, Hinton DR, Barbaro N, Annett GM, Raffel C. MDR1 gene expression in brain of patients with medically intractable epilepsy. *Epilepsia.* 1995;36(1):1-6.

[35] Ak H, Ay B, Tanriverdi T, Sanus GZ, Is M, Sar M, et al. Expression and cellular distribution of multidrug resistance-related proteins in patients with focal cortical dysplasia. *Seizure.* 2007;16(6):493-503.

[36] Aronica E, Gorter JA, Jansen GH, van Veelen CW, van Rijen PC, Leenstra S, et al. Expression and cellular distribution of multidrug transporter proteins in two major causes of medically intractable epilepsy: focal cortical dysplasia and glioneuronal tumors. *Neuroscience.* 2003;118(2):417-29.

[37] Aronica E, Gorter JA, Ramkema M, Redeker S, Ozbas-Gerceker F, van Vliet EA, et al. Expression and cellular distribution of multidrug resistance-related proteins in the hippocampus of patients with mesial temporal lobe epilepsy. *Epilepsia.* 2004;45(5):441-51.

[38] Dombrowski SM, Desai SY, Marroni M, Cucullo L, Goodrich K, Bingaman W, et al. Overexpression of multiple drug resistance genes in endothelial cells from patients with refractory epilepsy. *Epilepsia.* 2001;42(12):1501-6.

- [39] Kubota H, Ishihara H, Langmann T, Schmitz G, Stieger B, Wieser HG, et al. Distribution and functional activity of P-glycoprotein and multidrug resistance-associated proteins in human brain microvascular endothelial cells in hippocampal sclerosis. *Epilepsy Res.* 2006;68(3):213-28.
- [40] Lazarowski A, Lubieniecki F, Camarero S, Pomata H, Bartuluchi M, Sevlever G, et al. Multidrug resistance proteins in tuberous sclerosis and refractory epilepsy. *Pediatr Neurol.* 2004;30(2):102-6.
- [41] van Vliet EA, Zibell G, Pekcec A, Schlichtiger J, Edelbroek PM, Holtman L, et al. COX-2 inhibition controls P-glycoprotein expression and promotes brain delivery of phenytoin in chronic epileptic rats. *Neuropharmacology.* 2010;58(2):404-12.
- [42] Potschka H, Fedrowitz M, Loscher W. P-glycoprotein and multidrug resistance-associated protein are involved in the regulation of extracellular levels of the major antiepileptic drug carbamazepine in the brain. *Neuroreport.* 2001;12(16):3557-60.
- [43] Potschka H, Fedrowitz M, Loscher W. Multidrug resistance protein MRP2 contributes to blood-brain barrier function and restricts antiepileptic drug activity. *J Pharmacol Exp Ther.* 2003;306(1):124-31.
- [44] Clinckers R, Smolders I, Meurs A, Ebinger G, Michotte Y. Quantitative in vivo microdialysis study on the influence of multidrug transporters on the blood-brain barrier passage of oxcarbazepine: concomitant use of hippocampal monoamines as pharmacodynamic markers for the anticonvulsant activity. *J Pharmacol Exp Ther.* 2005;314(2):725-31.
- [45] van Vliet EA, van Schaik R, Edelbroek PM, Redeker S, Aronica E, Wadman WJ, et al. Inhibition of the multidrug transporter P-glycoprotein improves seizure control in phenytoin-treated chronic epileptic rats. *Epilepsia.* 2006;47(4):672-80.
- [46] Brandt C, Bethmann K, Gastens AM, Loscher W. The multidrug transporter hypothesis of drug resistance in epilepsy: Proof-of-principle in a rat model of temporal lobe epilepsy. *Neurobiol Dis.* 2006;24(1):202-11.
- [47] Hocht C, Lazarowski A, Gonzalez NN, Auzmendi J, Opezzo JA, Bramuglia GF, et al. Nimodipine restores the altered hippocampal phenytoin pharmacokinetics in a refractory epileptic model. *Neurosci Lett.* 2007;413(2):168-72.
- [48] Jambroszyk M, Tipold A, Potschka H. Add-on treatment with verapamil in pharmaco-resistant canine epilepsy. *Epilepsia.* 2011;52(2):284-91.

- [49] Sisodiya SM, Lin WR, Harding BN, Squier MV, Thom M. Drug resistance in epilepsy: expression of drug resistance proteins in common causes of refractory epilepsy. *Brain*. 2002;125(Pt 1):22-31.
- [50] Aronica E, Gorter JA, Redeker S, van Vliet EA, Ramkema M, Scheffer GL, et al. Localization of breast cancer resistance protein (BCRP) in microvessel endothelium of human control and epileptic brain. *Epilepsia*. 2005;46(6):849-57.
- [51] Lazarowski A, Massaro M, Schteinschnaider A, Intruvini S, Sevlever G, Rabinowicz A. Neuronal MDR-1 gene expression and persistent low levels of anticonvulsants in a child with refractory epilepsy. *Ther Drug Monit*. 2004;26(1):44-6.
- [52] Loscher W, Potschka H. Drug resistance in brain diseases and the role of drug efflux transporters. *Nat Rev Neurosci*. 2005;6(8):591-602.
- [53] Aronica E, Sisodiya SM, Gorter JA. Cerebral expression of drug transporters in epilepsy. *Adv Drug Deliv Rev*. 2012;64(10):919-29.
- [54] Sisodiya SM, Martinian L, Scheffer GL, van der Valk P, Scheper RJ, Harding BN, et al. Vascular colocalization of P-glycoprotein, multidrug-resistance associated protein 1, breast cancer resistance protein and major vault protein in human epileptogenic pathologies. *Neuropathol Appl Neurobiol*. 2006;32(1):51-63.
- [55] Calatuzzolo C, Gelati M, Ciusani E, Sciacca FL, Pollo B, Cajola L, et al. Expression of drug resistance proteins Pgp, MRP1, MRP3, MRP5 and GST-pi in human glioma. *J Neurooncol*. 2005;74(2):113-21.
- [56] Feldmann M, Asselin MC, Liu J, Wang S, McMahon A, Anton-Rodriguez J, et al. P-glycoprotein expression and function in patients with temporal lobe epilepsy: a case-control study. *Lancet Neurol*. 2013;12(8):777-85.
- [57] Feldmann M, Koepp M. P-glycoprotein imaging in temporal lobe epilepsy: in vivo PET experiments with the Pgp substrate [11C]-verapamil. *Epilepsia*. 2012;53 Suppl 6:60-3.
- [58] Marchi N, Guiso G, Rizzi M, Pirker S, Novak K, Czech T, et al. A pilot study on brain-to-plasma partition of 10,11-dihydro-10-hydroxy-5H-dibenzo(b,f)azepine-5-carboxamide and MDR1 brain expression in epilepsy patients not responding to oxcarbazepine. *Epilepsia*. 2005;46(10):1613-9.
- [59] Rambeck B, Jurgens UH, May TW, Pannek HW, Behne F, Ebner A, et al. Comparison of brain extracellular fluid, brain tissue, cerebrospinal fluid, and

serum concentrations of antiepileptic drugs measured intraoperatively in patients with intractable epilepsy. *Epilepsia*. 2006;47(4):681-94.

[60] Summers MA, Moore JL, McAuley JW. Use of verapamil as a potential P-glycoprotein inhibitor in a patient with refractory epilepsy. *Ann Pharmacother*. 2004;38(10):1631-4.

[61] Iannetti P, Spalice A, Parisi P. Calcium-channel blocker verapamil administration in prolonged and refractory status epilepticus. *Epilepsia*. 2005;46(6):967-9.

[62] Schmitt FC, Dehnicke C, Merschhemke M, Meencke HJ. Verapamil attenuates the malignant treatment course in recurrent status epilepticus. *Epilepsy Behav*. 2010;17(4):565-8.

[63] Pirker S, Baumgartner C. Termination of refractory focal status epilepticus by the P-glycoprotein inhibitor verapamil. *Eur J Neurol*. 2011;18(12):e151.

[64] Nicita F, Spalice A, Papetti L, Nikanorova M, Iannetti P, Parisi P. Efficacy of verapamil as an adjunctive treatment in children with drug-resistant epilepsy: a pilot study. *Seizure*. 2014;23(1):36-40.

[65] Borlot F, Wither RG, Ali A, Wu N, Verocai F, Andrade DM. A pilot double-blind trial using verapamil as adjuvant therapy for refractory seizures. *Epilepsy Res*. 2014;108(9):1642-51.

[66] Haerian BS, Lim KS, Tan CT, Raymond AA, Mohamed Z. Association of ABCB1 gene polymorphisms and their haplotypes with response to antiepileptic drugs: a systematic review and meta-analysis. *Pharmacogenomics*. 2011;12(5):713-25.

[67] Yu L, Liao WP, Yi YH, Qiu G. ABCB1 G2677T/A polymorphism is associated with the risk of drug-resistant epilepsy in Asians. *Epilepsy Res*. 2015;115:100-8.

[68] Lv WP, Han RF, Shu ZR. Associations between the C3435T polymorphism of the ABCB1 gene and drug resistance in epilepsy: a meta-analysis. *Int J Clin Exp Med*. 2014;7(11):3924-32.

[69] Li SX, Liu YY, Wang QB. ABCB1 gene C3435T polymorphism and drug resistance in epilepsy: evidence based on 8,604 subjects. *Med Sci Monit*. 2015;21:861-8.

[70] Zhang C, Kwan P, Zuo Z, Baum L. The transport of antiepileptic drugs by P-glycoprotein. *Adv Drug Deliv Rev*. 2012;64(10):930-42.

- [71] Luna-Tortos C, Fedrowitz M, Loscher W. Several major antiepileptic drugs are substrates for human P-glycoprotein. *Neuropharmacology*. 2008;55(8):1364-75.
- [72] Cerveny L, Pavek P, Malakova J, Staud F, Fendrich Z. Lack of interactions between breast cancer resistance protein (bcrp/abcg2) and selected antiepileptic agents. *Epilepsia*. 2006;47(3):461-8.
- [73] Luna-Tortos C, Fedrowitz M, Loscher W. Evaluation of transport of common antiepileptic drugs by human multidrug resistance-associated proteins (MRP1, 2 and 5) that are overexpressed in pharmaco-resistant epilepsy. *Neuropharmacology*. 2010;58(7):1019-32.
- [74] Romermann K, Helmer R, Loscher W. The antiepileptic drug lamotrigine is a substrate of mouse and human breast cancer resistance protein (ABCG2). *Neuropharmacology*. 2015;93:7-14.
- [75] Nakanishi H, Yonezawa A, Matsubara K, Yano I. Impact of P-glycoprotein and breast cancer resistance protein on the brain distribution of antiepileptic drugs in knockout mouse models. *Eur J Pharmacol*. 2013;710(1-3):20-8.
- [76] Shawahna R, Uchida Y, Decleves X, Ohtsuki S, Yousif S, Dauchy S, et al. Transcriptomic and quantitative proteomic analysis of transporters and drug metabolizing enzymes in freshly isolated human brain microvessels. *Mol Pharm*. 2011;8(4):1332-41.
- [77] Uchida Y, Ohtsuki S, Katsukura Y, Ikeda C, Suzuki T, Kamiie J, et al. Quantitative targeted absolute proteomics of human blood-brain barrier transporters and receptors. *J Neurochem*. 2011;117(2):333-45.
- [78] Englund G, Rorsman F, Ronnblom A, Karlbom U, Lazorova L, Grasjo J, et al. Regional levels of drug transporters along the human intestinal tract: co-expression of ABC and SLC transporters and comparison with Caco-2 cells. *Eur J Pharm Sci*. 2006;29(3-4):269-77.
- [79] Hilgendorf C, Ahlin G, Seithel A, Artursson P, Ungell AL, Karlsson J. Expression of thirty-six drug transporter genes in human intestine, liver, kidney, and organotypic cell lines. *Drug Metab Dispos*. 2007;35(8):1333-40.
- [80] Murakami T, Takano M. Intestinal efflux transporters and drug absorption. *Expert Opin Drug Metab Toxicol*. 2008;4(7):923-39.
- [81] Lin F, Marchetti S, Plum D, Iusuf D, Mazzanti R, Schellens JH, et al. Abcc4 together with abcb1 and abcg2 form a robust cooperative drug efflux system that

restricts the brain entry of camptothecin analogues. *Clin Cancer Res.* 2013;19(8):2084-95.

[82] Tiwari AK, Zhang R, Gallo JM. Overlapping functions of ABC transporters in topotecan disposition as determined in gene knockout mouse models. *Mol Cancer Ther.* 2013;12(7):1343-55.

[83] Vlaming ML, Pala Z, van Esch A, Wagenaar E, de Waart DR, van de Wetering K, et al. Functionally overlapping roles of Abcg2 (Bcrp1) and Abcc2 (Mrp2) in the elimination of methotrexate and its main toxic metabolite 7-hydroxymethotrexate in vivo. *Clin Cancer Res.* 2009;15(9):3084-93.

[84] Chen Y, Agarwal S, Shaik NM, Chen C, Yang Z, Elmquist WF. P-glycoprotein and breast cancer resistance protein influence brain distribution of dasatinib. *J Pharmacol Exp Ther.* 2009;330(3):956-63.

[85] Bauer M, Karch R, Zeitlinger M, Stanek J, Philippe C, Wadsak W, et al. Interaction of 11C-tarividar and 11C-elacridar with P-glycoprotein and breast cancer resistance protein at the human blood-brain barrier. *J Nucl Med.* 2013;54(8):1181-7.

[86] Agarwal S, Hartz AM, Elmquist WF, Bauer B. Breast cancer resistance protein and P-glycoprotein in brain cancer: two gatekeepers team up. *Curr Pharm Des.* 2011;17(26):2793-802.

[87] Polli JW, Olson KL, Chism JP, John-Williams LS, Yeager RL, Woodard SM, et al. An unexpected synergist role of P-glycoprotein and breast cancer resistance protein on the central nervous system penetration of the tyrosine kinase inhibitor lapatinib (N-{3-chloro-4-[(3-fluorobenzyl)oxy]phenyl}-6-[5-{{[2-(methylsulfonyl)ethyl]amino }methyl}-2-furyl]-4-quinazolinamine; GW572016). *Drug Metab Dispos.* 2009;37(2):439-42.

[88] de Vries NA, Zhao J, Kroon E, Buckle T, Beijnen JH, van Tellingen O. P-glycoprotein and breast cancer resistance protein: two dominant transporters working together in limiting the brain penetration of topotecan. *Clin Cancer Res.* 2007;13(21):6440-9.

[89] Potschka H, Luna-Munguia H. CNS transporters and drug delivery in epilepsy. *Curr Pharm Des.* 2014;20(10):1534-42.

- [90] Tiwari AK, Sodani K, Dai CL, Ashby CR, Jr., Chen ZS. Revisiting the ABCs of multidrug resistance in cancer chemotherapy. *Curr Pharm Biotechnol.* 2011;12(4):570-94.
- [91] Nicolle E, Boccard J, Guilet D, Dijoux-Franca MG, Zelefac F, Macalou S, et al. Breast cancer resistance protein (BCRP/ABCG2): new inhibitors and QSAR studies by a 3D linear solvation energy approach. *Eur J Pharm Sci.* 2009;38(1):39-46.
- [92] Henrich CJ, Robey RW, Bokesch HR, Bates SE, Shukla S, Ambudkar SV, et al. New inhibitors of ABCG2 identified by high-throughput screening. *Mol Cancer Ther.* 2007;6(12 Pt 1):3271-8.
- [93] Nicolle E, Boumendjel A, Macalou S, Genoux E, Ahmed-Belkacem A, Carrupt PA, et al. QSAR analysis and molecular modeling of ABCG2-specific inhibitors. *Adv Drug Deliv Rev.* 2009;61(1):34-46.
- [94] Matsson P, Englund G, Ahlin G, Bergstrom CA, Norinder U, Artursson P. A global drug inhibition pattern for the human ATP-binding cassette transporter breast cancer resistance protein (ABCG2). *J Pharmacol Exp Ther.* 2007;323(1):19-30.
- [95] Matsson P, Pedersen JM, Norinder U, Bergstrom CA, Artursson P. Identification of novel specific and general inhibitors of the three major human ATP-binding cassette transporters P-gp, BCRP and MRP2 among registered drugs. *Pharm Res.* 2009;26(8):1816-31.
- [96] Palmeira A, Rodrigues F, Sousa E, Pinto M, Vasconcelos MH, Fernandes MX. New uses for old drugs: pharmacophore-based screening for the discovery of P-glycoprotein inhibitors. *Chem Biol Drug Des.* 2011;78(1):57-72.
- [97] Falasca M, Linton KJ. Investigational ABC transporter inhibitors. *Expert Opin Investig Drugs.* 2012;21(5):657-66.
- [98] Jabeen I, Pleban K, Rinner U, Chiba P, Ecker GF. Structure-activity relationships, ligand efficiency, and lipophilic efficiency profiles of benzophenone-type inhibitors of the multidrug transporter P-glycoprotein. *J Med Chem.* 2012;55(7):3261-73.
- [99] Leong MK, Chen HB, Shih YH. Prediction of promiscuous p-glycoprotein inhibition using a novel machine learning scheme. *PLoS One.* 2012;7(3):e33829.
- [100] Pan Y, Chothe PP, Swaan PW. Identification of novel breast cancer resistance protein (BCRP) inhibitors by virtual screening. *Mol Pharm.* 2013;10(4):1236-48.

- [101] Tan KW, Li Y, Paxton JW, Birch NP, Scheepens A. Identification of novel dietary phytochemicals inhibiting the efflux transporter breast cancer resistance protein (BCRP/ABCG2). *Food Chem.* 2013;138(4):2267-74.
- [102] Zhang SL, Wei YX, Li Q, Sun HP, Peng H, You QD. Pharmacophore-based drug design and biological evaluation of novel ABCB1 inhibitors. *Chem Biol Drug Des.* 2013;81(3):349-58.
- [103] Chan HS, DeBoer G, Thiessen JJ, Budning A, Kingston JE, O'Brien JM, et al. Combining cyclosporin with chemotherapy controls intraocular retinoblastoma without requiring radiation. *Clin Cancer Res.* 1996;2(9):1499-508.
- [104] Deeken JF, Loscher W. The blood-brain barrier and cancer: transporters, treatment, and Trojan horses. *Clin Cancer Res.* 2007;13(6):1663-74.
- [105] Lhomme C, Joly F, Walker JL, Lissoni AA, Nicoletto MO, Manikhas GM, et al. Phase III study of valspodar (PSC 833) combined with paclitaxel and carboplatin compared with paclitaxel and carboplatin alone in patients with stage IV or suboptimally debulked stage III epithelial ovarian cancer or primary peritoneal cancer. *J Clin Oncol.* 2008;26(16):2674-82.
- [106] Teodori E, Dei S, Martelli C, Scapecchi S, Gualtieri F. The functions and structure of ABC transporters: implications for the design of new inhibitors of Pgp and MRP1 to control multidrug resistance (MDR). *Curr Drug Targets.* 2006;7(7):893-909.
- [107] Ding YL, Shih YH, Tsai FY, Leong MK. In silico prediction of inhibition of promiscuous breast cancer resistance protein (BCRP/ABCG2). *PLoS One.* 2014;9(3):e90689.
- [108] Shityakov S, Forster C. In silico structure-based screening of versatile P-glycoprotein inhibitors using polynomial empirical scoring functions. *Adv Appl Bioinform Chem.* 2014;7:1-9.
- [109] Montanari F, Ecker GF. BCRP Inhibition: from Data Collection to Ligand-Based Modeling. *Mol Inf.* 2014;33(5):322-31.
- [110] Cherigo L, Lopez D, Martinez-Luis S. Marine natural products as breast cancer resistance protein inhibitors. *Mar Drugs.* 2015;13(4):2010-29.
- [111] Mudududdla R, Guru SK, Wani A, Sharma S, Joshi P, Vishwakarma RA, et al. 3-(Benzo[d][1,3]dioxol-5-ylamino)-N-(4-fluorophenyl)thiophene-2-carboxamide

overcomes cancer chemoresistance via inhibition of angiogenesis and P-glycoprotein efflux pump activity. *Org Biomol Chem*. 2015;13(14):4296-309.

[112] Belekar V, Lingineni K, Garg P. Classification of Breast Cancer Resistant Protein (BCRP) Inhibitors and Non-Inhibitors Using Machine Learning Approaches. *Comb Chem High Throughput Screen*. 2015;18(5):476-85.

[113] Thai KM, Huynh NT, Ngo TD, Mai TT, Nguyen TH, Tran TD. Three- and four-class classification models for P-glycoprotein inhibitors using counter-propagation neural networks. *SAR QSAR Environ Res*. 2015;26(2):139-63.

[114] Callaghan R, Luk F, Bebawy M. Inhibition of the multidrug resistance P-glycoprotein: time for a change of strategy? *Drug Metab Dispos*. 2014;42(4):623-31.

[115] Cirrito JR, Deane R, Fagan AM, Spinner ML, Parsadanian M, Finn MB, et al. P-glycoprotein deficiency at the blood-brain barrier increases amyloid-beta deposition in an Alzheimer disease mouse model. *J Clin Invest*. 2005;115(11):3285-90.

[116] Bartels AL, Willemsen AT, Kortekaas R, de Jong BM, de Vries R, de Klerk O, et al. Decreased blood-brain barrier P-glycoprotein function in the progression of Parkinson's disease, PSP and MSA. *J Neural Transm (Vienna)*. 2008;115(7):1001-9.

[117] Vautier S, Fernandez C. ABCB1: the role in Parkinson's disease and pharmacokinetics of antiparkinsonian drugs. *Expert Opin Drug Metab Toxicol*. 2009;5(11):1349-58.

[118] Liu XD, Liu GQ. P glycoprotein regulated transport of glutamate at blood brain barrier. *Acta Pharmacol Sin*. 2001;22(2):111-6.

[119] Potschka H. Role of CNS efflux drug transporters in antiepileptic drug delivery: overcoming CNS efflux drug transport. *Adv Drug Deliv Rev*. 2012;64(10):943-52.

[120] Wu CP, Calcagno AM, Ambudkar SV. Reversal of ABC drug transporter-mediated multidrug resistance in cancer cells: evaluation of current strategies. *Curr Mol Pharmacol*. 2008;1(2):93-105.

[121] Potschka H. Targeting the brain--surmounting or bypassing the blood-brain barrier. *Handb Exp Pharmacol*. 2010(197):411-31.

[122] Rosillo-de la Torre A, Luna-Barcenas G, Orozco-Suarez S, Salgado-Ceballos H, Garcia P, Lazarowski A, et al. Pharmaco-resistant epilepsy and nanotechnology. *Front Biosci (Elite Ed)*. 2014;6:329-40.

[123] Bennewitz MF, Saltzman WM. Nanotechnology for delivery of drugs to the brain for epilepsy. *Neurotherapeutics*. 2009;6(2):323-36.

[124] Kapse-Mistry S, Govender T, Srivastava R, Yergeri M. Nanodrug delivery in reversing multidrug resistance in cancer cells. *Front Pharmacol*. 2014;5:159.

[125] European Medicines Agency. Guideline on the Investigations of Drug Interactions, (2012).

[126] Food and Drug Administration. Drug Interaction Studies: Study Design, Data Analysis, and Implications for Dosing, and Labeling Recommendations, (2012).



## Capítulo III

# Modelos QSAR

---



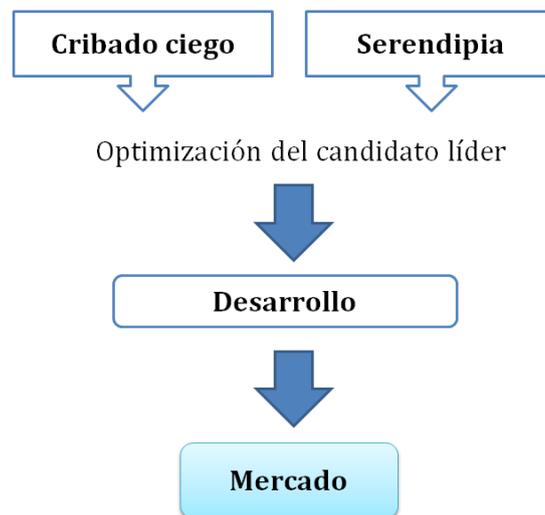
### ***3.1 Descubrimiento de fármacos: breve reseña histórica***

En las primeras civilizaciones los hombres confiaron en la naturaleza para sus necesidades de curación y los productos naturales han sido y siguen siendo la fuente más importante de fármacos o precursores de fármacos incluso en nuestros días [1-2]. El descubrimiento de fármacos ha sido originalmente un proceso fortuito y el conocimiento, generado a prueba y error, se fue compartiendo y conservando a través de las generaciones y mucho de ese conocimiento ha quedado registrado en la literatura [3-4]. Como usualmente se utilizaba la totalidad de una planta y/o alguna de sus partes, la mayoría de las personas terminaban por consumir múltiples entidades químicas al mismo tiempo, dado que aún se carecía del conocimiento de su contenido químico y de los efectos que este podía tener en el organismo [4].

Durante los primeros años del siglo XIX se introdujeron diversas técnicas que permitieron separar los componentes individuales de los extractos, lo que permitió allanar el camino para comenzar a utilizar entidades químicas individuales como medicamentos. Un ejemplo típico es la quinina, que si bien fue encontrada anteriormente, recién se aisló en 1823 [5]. Hasta finales del siglo XIX, el foco de la investigación estaba puesto principalmente en la prueba y evaluación de los productos naturales existentes [4]. Hacia finales de 1800 se establecieron algunas compañías químicas/farmacéuticas importantes como Bayer y Hoechst, y en 1895 los científicos de Bayer tuvieron éxito en la síntesis química de aspirina [6].

Los inicios de la industria farmacéutica moderna pueden situarse a principios del siglo XX [6], período en el que comenzó por ejemplo, el desarrollo de análogos de la quinina como medicamentos antimaláricos [5]. Se logró la síntesis de diversos fármacos gracias al avance y las mejoras de distintos procesos químicos; este concepto de fármacos sintéticos cobró mayor impulso en 1930 tras el descubrimiento de las sulfamidas. Por otro lado, gracias al impacto de los métodos de la biotecnología se consiguió la fabricación de insulina a gran escala. Como consecuencia de todo este desarrollo, en 1930 se estableció la Administración de Alimentos y Medicamentos de los Estados Unidos (*Food and Drug Administration*,

FDA), una autoridad reguladora cuyo principal objetivo era (en un primer momento) el de "proteger la salud pública". El descubrimiento de la penicilina por Alexander Fleming en 1928 fue seguido durante la década de 1940 por el descubrimiento de varias otras moléculas terapéuticas que derivaban de microorganismos [7]. Al mismo tiempo, tuvo lugar un importante desarrollo de otros fármacos como antihistamínicos, analgésicos, barbitúricos, hormonas, sedantes, hipnóticos y antidepresivos. A pesar de tener un origen biológico, varias de estas moléculas fueron obtenidas por síntesis química directa [6, 8]. En este período también se aislaron numerosos medicamentos de origen natural y se desarrollaron varios fármacos sintéticos [4]. La búsqueda de productos activos mediante el estudio de la medicina tradicional tiene una base más racional, aunque puede suponerse que originalmente fueron identificados mediante hallazgos fortuitos. El descubrimiento de fármacos mediante cribado sistemático y serendipia prosiguió hasta mediados del siglo XX, conduciendo al descubrimiento de diversos fármacos de éxito como la clorpromazina, meprobamato y las benzodiazepinas [3]. La Figura 3.1 representa el proceso de descubrimiento de fármacos seguido por las empresas farmacéuticas a mediados del siglo XX.



**Figura 3.1.** Esquema del proceso de descubrimiento de fármacos a mediados del siglo XX. La serendipia y la detección ciega dominaron el proceso.

Sin embargo, la necesidad de batallar enfermedades concretas y algunos inconvenientes importantes como el elevado consumo de tiempo de los procesos, los altos costos y la disminución de la garantía de éxito probaron que la búsqueda convencional de fármacos de manera aleatoria ya no era efectiva [7]. Es así como a fines de 1960 surgió el concepto de "diseño racional de fármacos" que revolucionó sustancialmente el paradigma vigente hasta el momento. El concepto de relación estructura-actividad cuantitativa (*Quantitative Structure-Activity Relationship*, QSAR) introducido por Hansch y Fujita en 1960 [9] dio el puntapié inicial para el diseño de fármacos asistido por computadora (*Computer-Aided Drug Design*, CADD).

Luego de aproximadamente diez años, comenzó a utilizarse intensamente el conocimiento producido por la biología estructural en el descubrimiento de fármacos. Como resultado de la fusión de la biotecnología con la industria farmacéutica, durante las últimas tres décadas del siglo XX se dio el surgimiento de varias compañías biofarmacéuticas. Varios de los blancos moleculares identificados por esos años fueron de naturaleza proteica, en particular enzimas. En 1980, se introdujeron los ensayos *in vitro* utilizando tejidos animales como una alternativa previa y a la vez complementaria de los métodos *in vivo* convencionales más invasivos utilizados hasta el momento, lo que permitió reducir los costos de experimentación y mejorar la tasa final de éxito [3, 10].

Hasta 1990 los mayores avances se dieron en la síntesis, la evaluación biológica y el análisis de la relación estructura-actividad de nuevos compuestos. Esto contribuyó a la optimización de los líderes utilizando técnicas tradicionales de la química medicinal [8]. La "Regla de los Cinco" de Lipinski, concebida para descartar tempranamente candidatos activos con propiedades biofarmacéuticas desfavorables, fue uno de los nuevos conceptos que alteraron drásticamente el paradigma tradicional para el descubrimiento y desarrollo de fármacos [11].

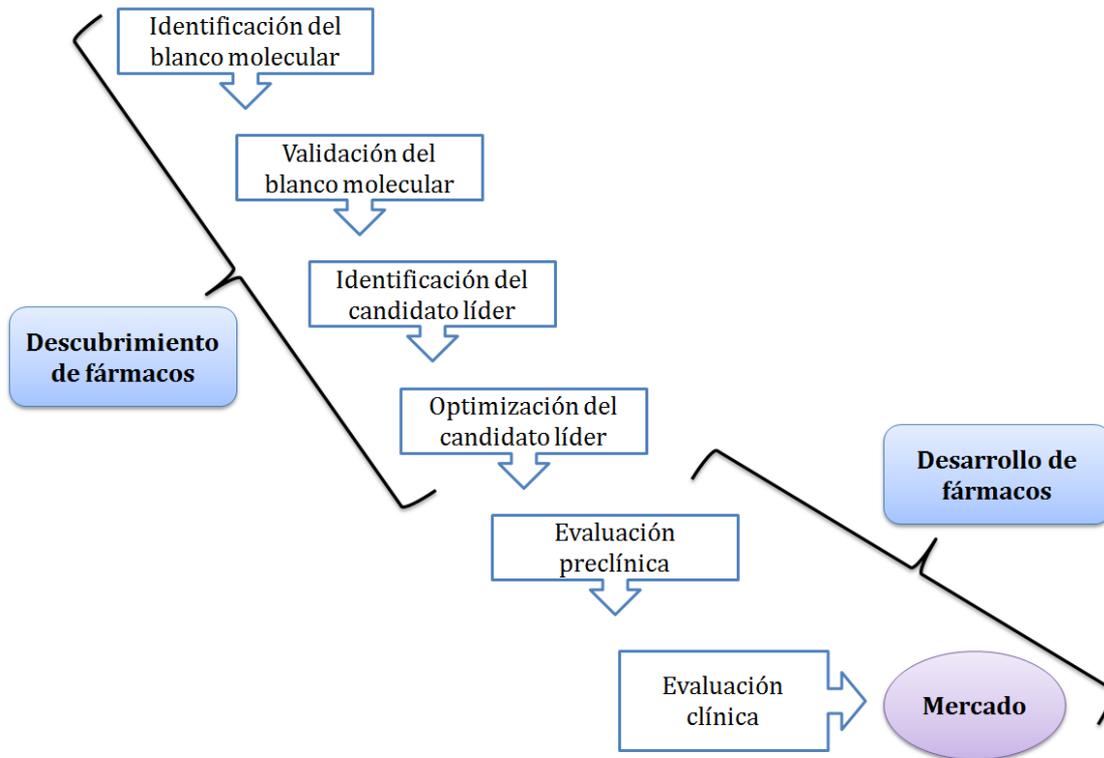
En las últimas dos décadas del siglo XX, grandes inversiones por parte de la industria farmacéutica llevaron al desarrollo de nuevas tecnologías y metodologías. Por un lado, el cribado de alto rendimiento (*High-throughput screening*, HTS) que posibilitó la evaluación de pequeñas muestras de una serie de compuestos contra una amplia gama de ensayos *in vitro* y permitió a los

investigadores identificar rápidamente aquellos candidatos prometedores de entre grandes bibliotecas de compuestos para un posterior desarrollo [12]. Por otro lado, la química combinatoria que permitió a los químicos sintetizar grandes cantidades de diversas bibliotecas de compuestos químicos optimizando costos y tiempos [13]. El proceso de descubrimiento de fármacos se transformó drásticamente debido al continuo desarrollo de la biología molecular (por ejemplo, las técnicas de clonación de genes, la tecnología de expresión *in vitro* y de mutagénesis dirigida al sitio, entre otras) [6, 8, 14].

El amanecer del siglo XXI fue testigo del inicio de la era post-genómica que encontró su máxima expresión en la finalización del Proyecto Genoma Humano (PGH) con un borrador inicial del genoma terminado en el año 2000 y el genoma completo presentado en abril de 2003 [15-16]. Uno de los mayores objetivos de este proyecto fue la identificación de blancos moleculares farmacológicos novedosos [17-18]. La necesidad de establecer en primer lugar si el blanco molecular identificado califica como farmacéuticamente explotable, es decir, si puede ser capaz de interactuar con un fármaco se llamó "farmacología inversa" [19]. La "revolución de las ómicas" como la genómica y la proteómica (en particular la tecnología de "*microarrays*" y de secuenciación) han permitido un nuevo cambio de paradigma en el desarrollo de nuevos fármacos [20].

En los últimos años, los investigadores que tradicionalmente se centraron en el desarrollo de un fármaco que actúe contra un blanco molecular determinado, han desplazado su atención a la modulación simultánea de múltiples objetivos en base al avance del conocimiento de la complejidad de las enfermedades [21]. Actualmente, el descubrimiento de fármacos se ha convertido en un campo científico interdisciplinario integrando diversas disciplinas de la biología, la química, las matemáticas y la computación [22]. Particularmente, la ciencias de la computación han cambiado profundamente la forma en la que "hacemos ciencia" o -mejor dicho- producimos conocimiento en todos los diferentes campos de investigación [23]. El descubrimiento y desarrollo de nuevos fármacos con mayor rapidez y a menor costo así como mejorar la eficacia sobre desórdenes complejos se han convertido en las necesidades principales del sector farmacéutico y los enfoques computacionales son fundamentales para lograrlo [24]. En la Figura 3.2

se muestra de manera esquemática el proceso de descubrimiento de fármacos actual.



**Figura 3.2.** Esquema del proceso de descubrimiento de fármacos moderno, que consta de siete pasos consecutivos. El proceso es largo y requiere de mucho tiempo. La aplicación de métodos computacionales en las primeras etapas ha reducido considerablemente los tiempos y costos implicados.

### 3.2 Métodos computacionales en el descubrimiento de fármacos

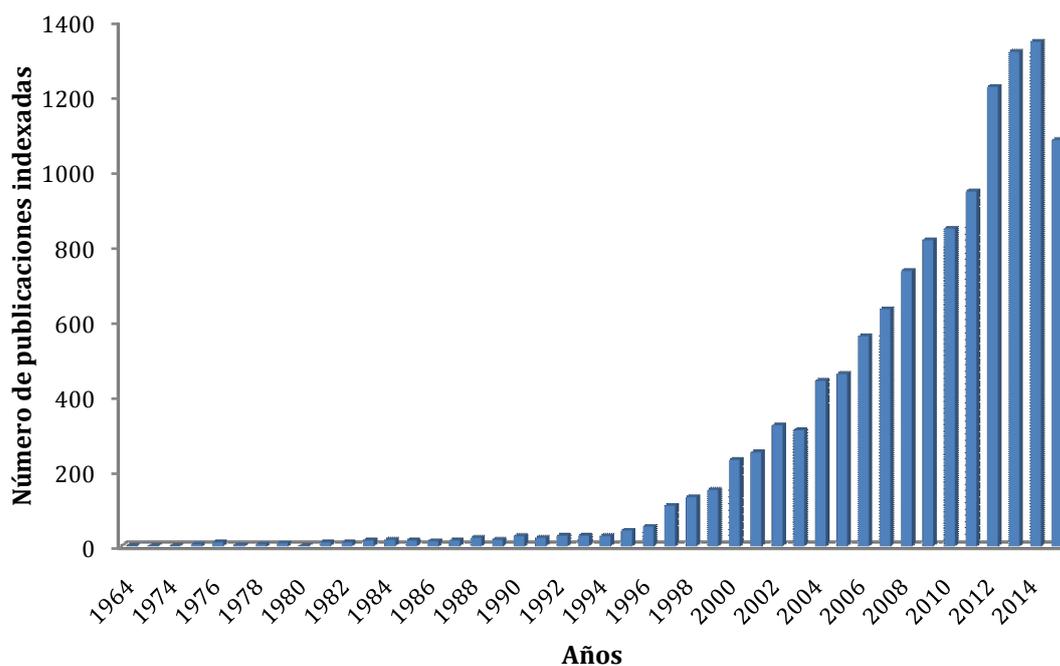
Se sabe comúnmente que el descubrimiento de nuevos fármacos es un proceso complejo, largo, arriesgado y costoso. Se estima que se gastan entre US\$ 0.3-1.8 billones en el desarrollo de un nuevo fármaco dependiendo del autor y de la indicación terapéutica que se esté analizando [25], mientras que el tiempo medio transcurrido hasta la aprobación del nuevo medicamento (sin considerar el tiempo que demanda identificar y validar un blanco molecular) es de

aproximadamente unos 13.5 años [26]. La introducción de soluciones de integración y gestión del conocimiento asistidas por computadora ha iniciado una nueva era en el descubrimiento de fármacos [27]. De hecho, las pasadas tres décadas han sido testigo del desarrollo de pequeñas moléculas terapéuticas exclusivamente en base a métodos de descubrimiento/diseño de fármacos asistido por computadora [28].

Hoy en día, la quimioinformática y la bioinformática se han convertido en grandes aliados para racionalizar el descubrimiento de nuevos fármacos. El CADD en comparación con el HTS y la química combinatoria tradicional realiza una búsqueda mejor orientada y ayuda a optimizar tiempos, recursos y costos [29-30].

El CADD aplicado al descubrimiento de fármacos se utiliza generalmente para tres propósitos principales: (1) filtrar grandes bibliotecas de compuestos químicos en conjuntos más pequeños predichos como potenciales activos para ser evaluados de forma experimental; (2) guiar la optimización de compuestos líderes, ya sea para aumentar su afinidad u optimizar su metabolismo y propiedades farmacocinéticas incluyendo la absorción, distribución, metabolismo, excreción, y el potencial de toxicidad (ADMET) y; (3) diseño de nuevos compuestos, ya sea por la incorporación de un grupo funcional a la vez a una molécula de partida o bien por el ensamblado de fragmentos para generar una nueva entidad química [28].

Particularmente, el propósito del punto 1 se conoce como cribado o tamizado virtual (*Virtual Screening, VS*), que consiste en un conjunto heterogéneo de técnicas computacionales que permiten explorar representaciones digitales de entidades químicas registradas en las bases de datos de compuestos químicos, con el fin de identificar posibles candidatos que reúnan una serie de requisitos o características estructurales que están de alguna manera vinculados a ciertas propiedades biológicas y/o químicas de interés [31]. Como se ilustra en la Figura 3.3, el uso del VS ha ido aumentando progresivamente desde las primeras campañas.



**Figura 3.3.** Evolución del número de publicaciones indexadas hasta el año 2015 en PubMed que se obtienen introduciendo “*virtual screening*” como palabra clave para la búsqueda.

### 3.2.1 Cribado Virtual (VS)

El esquema de clasificación más básica de las técnicas de VS las divide en:

- metodologías basadas en la estructura (o receptor) y,
- metodologías basadas en el ligando

Las metodologías basadas en la estructura pueden aplicarse siempre que esté disponible la estructura tridimensional del blanco molecular de interés (generalmente una proteína) a través de cristalografía de rayos X, Resonancia Magnética Nuclear (RMN) o esfuerzos de modelado comparativo o por homología [31]. Dentro de estas metodologías cabe mencionar al *docking* molecular que implica simular el acoplamiento y reconocimiento de cada uno de los compuestos de la base de datos a cribar (en una o más conformaciones) con el sitio activo del blanco molecular.

La mayoría de las estructuras de proteínas disponibles corresponden a proteínas solubles en medio acuoso, debido a las dificultades intrínsecas vinculadas a la obtención de proteínas de membrana/transmembrana tales como canales iónicos, los transportadores ABC, etc [32-33]. Las proteínas de membrana generalmente están presentes en bajos niveles en las membranas biológicas, por lo tanto no se pueden obtener fácilmente en las cantidades suficientes desde su entorno nativo y es necesario recurrir a su sobreexpresión en algún vector adecuado. Una dificultad importante para la sobreexpresión de proteínas de membrana en tales vectores es la agregación de la proteína en el citoplasma de la célula huésped y la ausencia de modificaciones post-traduccionales presentes en la célula original. En segundo lugar, las proteínas de membrana están embebidas en una bicapa lipídica, dinámica y compleja, lo que limita la aplicación de las técnicas biofísicas estándar para determinar su estructura. Por último, las proteínas de membrana son generalmente no solubles en solución acuosa, y requieren sistemas sintéticos especiales (por ejemplo, detergentes) para poder manipularlas *in vitro*. Las limitaciones para caracterizar las proteínas de membrana y el hecho de que se estima que alrededor del 30% del genoma humano codifica proteínas de membrana [34] explica la importancia de las técnicas basadas en el ligando cuando la estructura del blanco molecular no está disponible. Los enfoques basados en el receptor para el desarrollo de nuevos fármacos para eludir la BCRP aún no se han explorado extensamente debido justamente a la falta de una estructura tridimensional de alta resolución de esta proteína transportadora transmembrana, si bien se han realizado predicciones de la estructura tridimensional de la BCRP mediante el desarrollo de modelos por homología sobre la base de la similitud de las secuencias proteicas con estructuras altamente conocidas [35-36].

Debido a esta limitación, en la presente tesis se ha decidido aplicar las metodologías basadas en el ligando, que podemos agrupar en tres sub-categorías:

- enfoques basados en el farmacóforo,
- enfoques basados en descriptores moleculares y,
- enfoques basados en similitud molecular

La presente tesis se centra específicamente en los enfoques basados en descriptores moleculares. Un descriptor molecular es una variable numérica,

discreta o continua, que refleja distintos aspectos de la estructura molecular. Todeschini y Consonni brindan la definición más aceptada hasta el momento:

*“Los descriptores moleculares son el resultado final de un procedimiento lógico y matemático el cual transforma la información química codificada dentro de una representación simbólica de una molécula en un número útil, o el resultado de un experimento estandarizado” [37].*

Los descriptores moleculares pueden clasificarse en dos categorías principales (Figura 3.4):

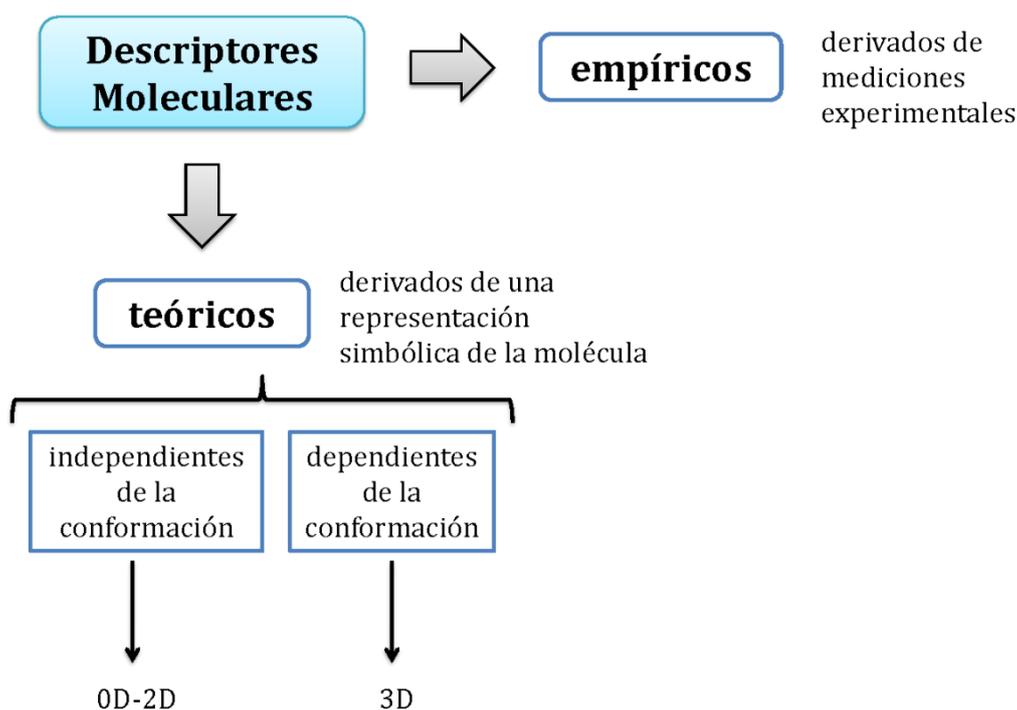
- Descriptores empíricos, derivados de mediciones experimentales y;
- Descriptores teóricos, derivados de una representación simbólica de la molécula.

Cada representación molecular constituye una forma diferente de mirar, interpretar y extraer información de la estructura de una molécula. Los descriptores teóricos son particularmente útiles en el campo del VS debido a que no requieren ninguna determinación experimental para caracterizar las moléculas de la base de datos sometida a cribado *in silico*. Esencialmente, los descriptores moleculares teóricos pueden clasificarse según la dimensionalidad de la representación molecular de la cual se derivan en:

a) Descriptores independientes de la conformación y/o de la orientación en el espacio de la molécula, donde el valor que asume el descriptor no depende del conformero de la estructura a partir de la cual se calculan ni de la orientación del mismo en un marco de referencia determinado. Puede llamárselos también descriptores de baja dimensionalidad (0D-2D). Este grupo engloba varios subgrupos de descriptores moleculares tales como los descriptores 0D y 1D, entre los cuales se encuentran los descriptores constitucionales, derivados de la fórmula química, tales como el peso molecular o el número de átomos de un elemento químico dado que están presentes en el compuesto; variables que señalan la presencia/ausencia o frecuencia de un tipo de átomo, grupo funcional o función química y asumen valores binarios (en el caso que indiquen ausencia o presencia) o en su defecto valores no binarios pero discretos (cuando indican la frecuencia de aparición en una estructura determinada); y los descriptores 2D (como los descriptores topológicos derivados de la Teoría de Grafos).

b) Descriptores dependientes de la conformación, cuyo valor sí depende de la conformación y/o en algunos casos de la orientación de la molécula en relación a un marco de referencia. Debido a la primera condición, la conformación debe ser, cuando menos, una conformación energética probable. Usualmente se definirá la conformación más probable como la conformación de mínima energía (esto es, el mínimo global de energía en el espacio conformacional, calculado con un método determinado). Encontrar tal conformación implica un muestreo sistemático o estocástico del espacio conformacional. Este tipo de descriptores suele denominarse descriptores 3D.

En este trabajo nos hemos decidido por el uso de descriptores de baja dimensionalidad, que no requieren un análisis conformacional de las estructuras de entrenamiento ni de las de la base de datos sometida posteriormente a cribado.

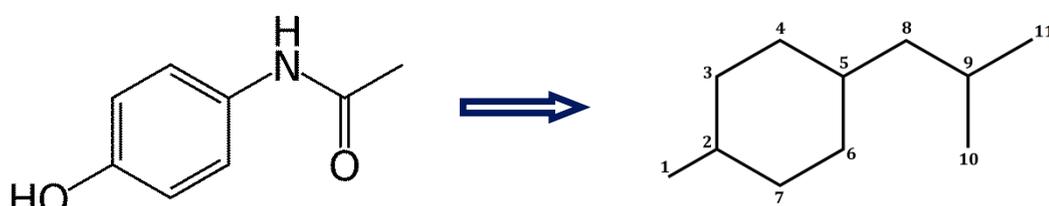


**Figura 3.4.** Clasificación de los descriptores moleculares.

Como ya mencionamos, los descriptores topológicos son descriptores derivados de una rama de las matemáticas conocida como la Teoría de Grafos que

fue inaugurada en 1736 por Euler [38]. La topología puede definirse como aquella parte del álgebra que estudia las posiciones e interconexiones de los elementos dentro de un conjunto. Aplicada a las moléculas, ha dado origen a una nueva disciplina llamada “topología molecular” o Teoría de Grafos Química, que analiza las posiciones relativas y la conectividad de los átomos dentro de las moléculas. Desde este punto de vista no se aborda el estudio de aspectos tales como la estructura tridimensional del compuesto, los tipos de enlaces, ángulos entre ellos y, a veces, ni siquiera la naturaleza de los átomos enlazados. La cuestión fundamental recae en qué átomos están ligados entre sí y cuáles caminos conducen de un átomo a otro dentro de la molécula.

El método topológico emplea los llamados “índices topológicos” para caracterizar estructuralmente a un compuesto. Para definir dichos índices, el primer paso es representar a los átomos por puntos (vértices) y a los enlaces por segmentos (ejes o aristas), eliminándose ocasionalmente los átomos de hidrógeno. De esta manera, se obtiene el “grafo” de la molécula (Figura 3.5). Posteriormente, se numeran aleatoriamente los distintos átomos (o vértices) y se construye alguna “matriz topológica” (por ejemplo, matriz distancia, matriz adyacencia, matriz de distancias topológicas máximas). A partir de esta matriz y mediante operaciones algebraicas más o menos sencillas se obtienen los “índices topológicos” más importantes, que pueden caracterizar aspectos muy diversos de la estructura molecular tales como la forma, el grado de ramificación o ciclación de una estructura química o la distribución de cargas.



**Figura 3.5.** Grafo asociado a la estructura química del paracetamol. Los índices topológicos son invariantes del grafo y, por lo tanto, es indistinto dónde se inicie la numeración del grafo.

### ***3.3 Modelos de relación estructura-actividad cuantitativa***

Los descriptores moleculares son útiles para obtener relaciones cuantitativas estructura-actividad/propiedad (QSAR/QSPR). El principio subyacente a las relaciones QSAR es que estructuras moleculares diferentes presentarán propiedades moleculares diferentes y estructuras moleculares similares presentarán propiedades similares (Principio de Similitud de Johnson y Maggiora) [39]. Bajo esta premisa, aplicando aproximaciones estadísticas se logra obtener correlaciones significativas entre una variable dependiente (la propiedad o actividad de un conjunto de compuestos de interés) y uno o más descriptores moleculares. La fórmula matemática más general de un modelo QSAR se representa en la siguiente ecuación:

$$\text{Actividad} = f(\text{propiedades fisicoquímicas y/o propiedades estructurales})$$

La principal aplicación de este tipo de relaciones matemáticas entre las características estructurales y una propiedad química o biológica de un conjunto de compuestos es la predicción de dicha propiedad para un conjunto de compuestos antes de (o incluso en lugar de) sintetizarlos y/o medir dicha propiedad/actividad experimentalmente, lo cual suele ser muy costoso [40].

#### ***3.3.1 QSAR desde una perspectiva histórica***

La raíz del enorme árbol de la Teoría QSAR se puede encontrar en la tesis de Cros del año 1863 (Facultad de Medicina de la Universidad de Estrasburgo, Francia), quien advirtió la relación entre la toxicidad de alcoholes primarios alifáticos y su solubilidad en agua. Sin duda, la focalización en el concepto de "estructura molecular" que se dio entre los años 1860 y 1880 allanó el camino para el ulterior desarrollo de la Teoría QSAR [41]. Así, Brown y Fraser [42] propusieron la existencia de una correlación entre la constitución molecular y la actividad biológica de diferentes alcaloides, mientras que algo más tarde Mills observaba las

reacciones entre la composición química y propiedades fisicoquímicas tales como el punto de fusión y el punto de ebullición [43]. El trabajo desarrollado por Hammett entre los años 1935 y 1938 sobre los efectos de los sustituyentes en las reacciones orgánicas se consideran un hito en el desarrollo de los modelos QSAR, en tanto suponen un importante avance desde las relaciones estructura-propiedad cualitativas hacia relaciones cuantitativas [44-47]. Algunos años más tarde se propusieron los primeros estudios teóricos QSAR/QSPR basados en los descriptores obtenidos a partir de la Teoría de Grafos [48-49], a lo que sobrevino el desarrollo de otras categorías de descriptores moleculares [50-53]. El año 1962 suele considerarse el inicio de la Teoría QSAR moderna; entonces, Hansch y col. correlacionaron la actividad biológica de reguladores del crecimiento vegetal y derivados de cloromicetina con las constantes de Hammett y los coeficientes de reparto [54]. Su trabajo fue el primer modelo QSAR multi-paramétrico. A continuación otros trabajos de Hansch y col. [55-56] atrajeron gran interés en este campo y condujeron a la explosión del QSAR [57].

Un hecho histórico importante en los estudios de estructura-actividad fue la publicación de Free y Wilson en 1964 [58], donde la idea básica del enfoque propuesto era el modelado de una actividad biológica (o propiedad química) mirando la presencia o ausencia de grupos funcionales sustituyentes en un esqueleto estructural común.

En la década de 1980 se dio el ingreso de la geometría 3D de las moléculas en el QSAR/QSPR [41] lo que derivó en un aumento en la complejidad del método y en su capacidad de descripción. En esta década, otra etapa brillante y realmente importante para el QSAR fue el desarrollo de descriptores moleculares basados en campos de interacción molecular (*molecular interaction fields*, MIFs) que dieron lugar al bien conocido campo del QSAR 3D [59-60]. Históricamente, el primer enfoque 3D QSAR fue el método GRID propuesto por Goodford en 1985 [61] y luego desarrollado por Cramer y col. en 1988 con el nombre de análisis comparativo de campos moleculares (*comparative molecular field analysis*, CoMFA) [62]. Otros métodos 3D QSAR importantes basados en MIF fueron introducidos con posterioridad [63-67].

Hoy en día el QSAR/QSPR se presenta como una herramienta de gran utilidad en el diseño molecular para diferentes propósitos [68-69] y resulta esencial e inseparable para el descubrimiento de nuevos fármacos y la optimización de la actividad durante su proceso de desarrollo [60, 70-73].

### 3.3.2 Construcción de un modelo QSAR

Un modelo QSAR se construye a través de las siguientes etapas generales [74]:

- 1- Selección o construcción de una base de datos que contenga un subconjunto de compuestos químicos o biológicos con y sin la actividad o propiedad de interés,
- 2- Cálculo de los descriptores moleculares,
- 3- Selección de los descriptores adecuados en el contexto de la actividad analizada para incluir en el modelo,
- 4- Búsqueda y construcción del modelo matemático óptimo que correlacione una actividad biológica o propiedad específica con la estructura molecular y,
- 5- Validación adecuada de dichos modelos.

Dentro de los métodos de selección de variables (en este caso, descriptores moleculares) podemos mencionar dos grupos, por un lado los métodos lineales que se basan en el supuesto de la existencia de una relación lineal entre las variables independientes y la variable dependiente o respuesta, y por otro lado los enfoques no lineales, donde los descriptores se asignan principalmente a un espacio relacional no lineal y ayudan a superar algunas limitaciones de los métodos lineales [75-78]. En la Tabla 3.1 se detallan algunos de los métodos más usados para la selección de variables tanto lineales como no lineales. En todos los métodos las variables se introducen en el modelo a través de una forma algorítmica y una función de aptitud (o de "fitness") o criterios de selección determinan qué variable debe ser retenida o eliminada del modelo. Se han introducido en la literatura diferentes criterios de selección [79-81], pero utilizar un criterio de selección depende de la finalidad del modelo QSAR/QSPR. Si el

objetivo es la predicción, el criterio que se utiliza es evaluar la capacidad de predicción del modelo generado.

**Tabla 3.1.** Listado de los métodos más conocidos de selección de variables tanto lineales como no lineales.

Métodos lineales	Métodos no lineales
Selección por pasos hacia adelante (FS)	Algoritmo genético (GA)
Eliminación por pasos hacia atrás (BE)	Programación evolutiva (EP)
Selección por pasos (combina los dos métodos anteriores)	Recocido simulado generalizado (GSA)
Método de sustitución (RM)	Sistema de colonia de hormigas (ACS)
Método de sustitución mejorado (ERM)	Optimización de enjambre de partículas (PSO)

La correlación entre la propiedad química o biológica de interés y la estructura molecular se basará en métodos de regresión/ajuste que ponderan la contribución de cada una de las variables seleccionadas a la propiedad modelada. Estos métodos también pueden ser lineales y no lineales. La elección del tipo de método a emplear varía en cada caso y no suele ser evidente en primera instancia dado que ninguna técnica es consistentemente mejor que todas las demás. Por regla general, se prefiere ir de los métodos más sencillos a los más complejos [40, 82].

Algunos de los métodos que se utilizan para la construcción de los modelos independientes después de la etapa de selección de variables o que incluyen una estrategia de selección de variables son [83]:

- Regresión lineal múltiple (MLR)
- Regresión de componentes principales (PCR)
- Mínimos cuadrados parciales (PLS)
- Red neuronal artificial (ANN)
- Máquina de soporte vectorial para regresión (SVMR)
- Árboles de decisión (DT)

Finalmente las predicciones dependen de la fiabilidad del modelo QSAR desarrollado. Se ha demostrado ampliamente que el ajuste de los datos al modelo establecido (sin ningún tipo de restricción en la construcción del modelo) no es evidencia de su validez y fiabilidad [84-86]. La validación adecuada del modelo es uno de los pasos fundamentales. Se pueden mencionar cuatro herramientas de validación que son las más conocidas y utilizadas [40]:

- 1- validación cruzada,
- 2- aleatorización de la variable dependiente,
- 3- remuestreo o "*bootstrapping*" y,
- 4- validación externa mediante un conjunto de compuestos de prueba de actividad conocida.

En el siguiente capítulo se profundizará específicamente en cada una de las metodologías aplicadas en la presente tesis.

### ***3.4 QSAR aplicado al cribado ADMET in silico***

A pesar de los constantes avances en el conocimiento de las bases moleculares de las enfermedades y aunque la inversión de la industria farmacéutica en I + D ha crecido de manera constante [87] el sector farmacéutico se enfrenta a una crisis de productividad. Las investigaciones sobre las causas de los fracasos en la última etapa del desarrollo de fármacos, realizada en la década de 1990, revelaron que las propiedades farmacocinéticas pobres y la elevada toxicidad fueron a menudo las responsables [88-89]. Hoy día, entre los candidatos que llegan a los ensayos clínicos, las tasas de deserción más altas se observan en las fases clínicas tardías (es decir, de fase II y III) sobre todo debido a la toxicidad o la falta de eficacia [25-26, 90-91]. Lo que es aún más significativo, es que las tasas de deserción en todas las fases clínicas han ido aumentando de manera constante a través de los años [26, 92-93].

A comienzos de 1990, las características biofarmacéuticas/farmacocinéticas desfavorables fueron responsables de casi el 40% de los fallos; mientras que para el año 2000, gracias al reconocimiento temprano de potenciales fallas relacionadas

con propiedades ADMET, las tasas de fracasos vinculados a este problema se redujeron radicalmente y pasaron a contribuir en menos del 10% a la tasa de deserción mundial [94]. Por esta situación, los expertos tanto de los sectores público como privado parecen coincidir en que la detección temprana de potenciales fallas en los ensayos clínicos es una estrategia clave para mejorar el éxito global y la productividad [26, 93, 95-97] reduciendo consecuentemente el tiempo y los gastos del desarrollo [88-89, 98], lo que a menudo se refiere como el paradigma "fallar temprano, fallar barato".

En este contexto, los organismos reguladores más importantes del mundo han reconocido el papel clave que las tecnologías de química y bioinformática podrían tener en este asunto [99]. En particular, la metodología QSAR no sólo permite predecir la actividad de nuevos compuestos químicos diseñados sino que también permite eliminar aquellos candidatos con pobres propiedades farmacocinéticas o aquellos con alta probabilidad de respuesta tóxica de un mayor desarrollo farmacéutico [11, 82, 100].

En la presente tesis se aplicará la Teoría QSAR para el desarrollo de modelos computacionales que sean capaces de predecir la interacción de nuevos candidatos a fármacos con la BCRP y de esa manera reducir las chances de avanzar a futuras instancias de investigación sobre compuestos que tendrán altas probabilidades de presentar problemas de MDR y de interacciones medicamentosas mediadas por dicho transportador.

---

### ***Referencias Capítulo III***

- [1] Verpoorte R. Exploration of nature's chemodiversity: the role of secondary metabolites as leads in drug development. *Drug Discovery Today*. 1998;3(5):232-8.
- [2] Harvey AL. Natural products in drug discovery. *Drug Discov Today*. 2008;13(19-20):894-901.
- [3] Ratti E, Trist D. Continuing evolution of the drug discovery process. *Pure Appl Chem*. 2001;73(1):67-75.
- [4] Sneader W. *Drug Discovery: A history*. New York: John Wiley & Sons, Ltd; 2005.
- [5] Burger A. *Medicinal Chemistry*. New York: Wiley-Interscience; 1970.
- [6] Walsh G. *Pharmaceutical biotechnology: Concepts and applications*. United Kingdom: John Wiley; 2007.
- [7] Umashankar V, Gurunathan S. Drug discovery: an appraisal. *International Journal of Pharmacy and Pharmaceutical Sciences*. 2015;7(4):59-66.
- [8] Smith CG, O'Donnell JT. Overview of the current process of new drug discovery and development. In: Smith CG, O'Donnell JT, editors. *The process of new drug discovery and development*. United States: Informa Healthcare; 2006. p. 7-13.
- [9] Hansch C, Fujita T.  $p$ - $\sigma$ - $\pi$  analysis. A method for the correlation of biological activity and chemical structure. *J Am Chem Soc*. 1964;86(8):1616-26.
- [10] Congreve M, Murray CW, Blundell TL. Structural biology and drug discovery. *Drug Discov Today*. 2005;10(13):895-907.
- [11] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 2001;46(1-3):3-26.
- [12] Pereira DA, Williams JA. Origin and evolution of high throughput screening. *Br J Pharmacol*. 2007;152(1):53-61.
- [13] Abraham DJ. *Burger's medicinal chemistry and drug discovery*. 6th ed. Hoboken, New Jersey: John Wiley and Sons, Inc.; 2003.
- [14] Boehm HJ, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, et al. 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(14):2664-74.

[15] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(6822):860-921.

[16] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science.* 2001;291(5507):1304-51.

[17] Broder S, Venter JC. Sequencing the entire genomes of free-living organisms: the foundation of pharmacology in the new millennium. *Annu Rev Pharmacol Toxicol.* 2000;40:97-132.

[18] Ward SJ. Drug discovery and genomics technologies: Impact of genomics in drug discovery. *BioTechniques.* 2001;31(3):626-30.

[19] Day M, Rutkowski JL, Feuerstein GZ. Translational medicine--a paradigm shift in modern drug discovery and development: the role of biomarkers. *Adv Exp Med Biol.* 2009;655:1-12.

[20] Harikumar SL, Neha S. Use of genomics and proteomics in pharmaceutical drug discovery and development: a review. *International Journal of Pharmacy and Pharmaceutical Sciences.* 2013;5(3):24-8.

[21] Masoudi-Nejad A, Mousavian Z, Bozorgmehr JH. Drug-target and disease networks: polypharmacology in the post-genomic era. *In Silico Pharmacol.* 2013;1:17.

[22] Herrling PL. The drug discovery process. *Prog Drug Res.* 2005;62:1-14.

[23] Vallverdú J. Computational Epistemology and e-Science. A New Way of Thinking. *Minds and Machines.* 2009;19(4):557-67.

[24] Rahman MM, Karim MR, Ahsan MQ, Khalipha ABR, Chowdhury MR, Saifuzzaman M. Use of computer in drug design and drug discovery: A review. *Int J Pharm Life Sci.* 2012;1(2):1-21.

[25] Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D. The cost of drug development: a systematic review. *Health Policy.* 2011;100(1):4-17.

[26] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov.* 2010;9(3):203-14.

- [27] Baldi A. Computational approaches for drug design and discovery: an overview. . Syst Rev Pharm 2010;1:99-105.
- [28] Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational methods in drug discovery. Pharmacol Rev. 2014;66:334-95.
- [29] Saha D, Emran TB, Saha SP. Bioinformatics: The effects on the cost of drug discovery. . Galle Med J 2013;18(1):44-50.
- [30] Suryawanshi SB, Nazeruddin GM. Computer aided drug discovery and development - an important need of the hour. . Int J Pharm Chem Sci 2013;2(1):293-6.
- [31] Talevi A, Castro EA, Bruno-Blanch LE. Virtual Screening: An Emergent, Key Methodology for Drug Development in an Emergent Continent. A Bridge Towards Patentability. In: Castro E, Haghi A, editors. Advanced Methods and Applications in Chemoinformatics: Research Progress and New Applications. Hershey: PA: Engineering Science Reference; 2012. p. 229-45.
- [32] Lin SH, Guidotti G. Purification of membrane proteins. Methods Enzymol. 2009;463:619-29.
- [33] Smith SM. Strategies for the purification of membrane proteins. Methods Mol Biol. 2011;681:485-96.
- [34] Terstappen GC, Reggiani A. In silico research in drug discovery. Trends Pharmacol Sci. 2001;22(1):23-6.
- [35] Hazai E, Bikadi Z. Homology modeling of breast cancer resistance protein (ABCG2). J Struct Biol. 2008;162(1):63-74.
- [36] Li YF, Polgar O, Okada M, Esser L, Bates SE, Xia D. Towards understanding the mechanism of action of the multidrug resistance-linked half-ABC transporter ABCG2: a molecular modeling study. J Mol Graph Model. 2007;25(6):837-51.
- [37] Todeschini R, Consonni V. Handbook of Molecular Descriptors: Wiley-VCH; 2000.
- [38] Biggs N, Lloyd E, Wilson R. Graph Theory: Oxford University Press; 1986.
- [39] Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. Mol Inf. 2010;29:476-88.

- [40] Yousefinejad S, Hemmateenejad B. Chemometrics tools in QSAR/QSPR studies: A historical perspective. *Chemometrics and Intelligent Laboratory Systems*. 2015;149(PART B):177–204.
- [41] Todeschini R, Consonni V, Gramatica P. Chemometrics in QSAR. In: Tauler R, Walczak B, Brown SD, editors. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*. Amsterdam: Elsevier; 2009. p. 140–1.
- [42] Brown AC, Fraser TR. On the Connection between Chemical Constitution and Physiological Action; with special reference to the Physiological Action of the Salts of the Ammonium Bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. *J Anat Physiol*. 1868;2(2):224-42.
- [43] Mills EJ. XXIII. On melting-point and boiling-point as related to chemical composition. *Philos Mag Ser*. 1884;5(17):173-87.
- [44] Hammett LP. Reaction rates and indicator acidities. *Chem Rev* 1935;16:67–79.
- [45] Hammett LP. Some relations between reaction rates and equilibrium constants. *Chem Rev*. 1935;17:125–36.
- [46] Hammett LP. The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J Am Chem Soc*. 1937;59:96-103.
- [47] Hammett LP. Linear free energy relationships in rate and equilibrium phenomena. *Trans Faraday Soc*. 1938;34:156.
- [48] Platt JR. Influence of neighbor bonds on additive bond properties in paraffins. *J Chem Phys*. 1947;15:419.
- [49] Wiener H. Influence of interatomic forces on paraffin properties. *J Chem Phys*. 1947;15:766.
- [50] Fukui K, Yonezawa T, Nagata C. Theory of substitution in conjugated molecules. *Bull Chem Soc Jpn*. 1954;27:423–7.
- [51] Mulliken RS. Electronic population analysis on LCAO-MO molecular wave functions. I. *J Chem Phys*. 1955;23:1833.
- [52] Mulliken RS. Electronic population analysis on LCAO-MO molecular wave functions. II. Overlap populations, bond orders, and covalent bond energies. *J Chem Phys*. 1955;23:1841.
- [53] Kier LB. *Molecular Orbital Theory in drug Research*: Academic Press; 1971.

- [54] Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature*. 1962;194:178–80.
- [55] Hansch C, Muir RRMR, Fujita T, Maloney PP, Geiger F, Streich M. The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. *J Am Chem Soc*. 1963;85:2817–24.
- [56] Fujita T, Iwasa J, Hansch C. A new substituent constant,  $\pi$ , derived from partition coefficients. *J Am Chem Soc*. 1964;86:5175–80.
- [57] Hansch C, Leo A. *Exploring QSAR, Fundamentals and Applications in Chemistry and Biology*. Washington, DC: American Chemical Society; 1995.
- [58] Free SM, Wilson JW. A mathematical contribution to structure–activity studies. *J Med Chem*. 1964;7:395–9.
- [59] Kubinyi H. *3D QSAR in drug design, Theory Methods and Applications*. Dordrecht: Kluwer Academic Publishers; 2000.
- [60] Kubinyi H, Folkers G, Martin YC. *3D QSAR in drug design, Recent Advances*. New York: Kluwer Academic Publishers; 2002.
- [61] Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem*. 1985;28(7):849-57.
- [62] Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc*. 1988;110(18):5959-67.
- [63] Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem*. 1994;37(24):4130-46.
- [64] Jain AN, Koile K, Chapman D. Compass: predicting biological activities from molecular surface properties. Performance comparisons on a steroid benchmark. *J Med Chem*. 1994;37(15):2315-27.
- [65] Silverman BD, Platt DE. Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition. *J Med Chem*. 1996;39(11):2129-40.

- [66] Chuman H, Karasawa M, Fujita T, Kansai F. QSAR a novel three-dimensional QSAR procedure: Voronoi field analysis. *Quant Struct-Act Relat.* 1998;17(4):313–26.
- [67] Cruciani G, Pastor M, Guba W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur J Pharm Sci.* 2000;11 Suppl 2:S29-39.
- [68] Khan MT. Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. *Curr Drug Metab.* 2010;11(4):285-95.
- [69] Quintero FA, Patel SJ, Muñoz F, Sam Mannan M. Review of existing QSAR/QSPR models developed for properties used in hazardous chemicals classification system. *Ind Eng Chem Res* 2012;51(49):16101–15.
- [70] Kubinyi H. QSAR and 3D QSAR in drug design part 2: applications and problems. *Drug Discov Today.* 1997;2(12):538–46.
- [71] Hopfinger AJ. Practical applications of computer-aided drug design. In: Charifson PS, editor. *Practical Application of Computer-Aided Drug Design.* New York: Marcel-Dekker; 1997. p. 105–64.
- [72] Perkins R, Fang H, Tong W, Welsh WJ. Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. *Environ Toxicol Chem.* 2003;22(8):1666-79.
- [73] Du QS, Huang RB, Chou KC. Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. *Curr Protein Pept Sci.* 2008;9(3):248-60.
- [74] Aguiar-Pulido V, Gestal M, Cruz-Monteagudo M, Rabunal JR, Dorado J, Munteanu CR. Evolutionary computation and QSAR research. *Curr Comput Aided Drug Des.* 2013;9(2):206-25.
- [75] Goodarzi M, Deshpande S, Murugesan V, Katti S, Prabhakar Y. Is feature selection essential for ANN modeling?. *QSAR Comb Sci.* 2009;28(11-12):1487–99.
- [76] Goodarzi M, Dejaegher B, Vander Heyden Y. Feature selection methods in QSAR studies. *J AOAC Int.* 2012;95(3):636-51.
- [77] Goodarzi M, Vander Heyden Y, Funar-Timofei S. Towards better understanding of feature-selection or reduction techniques for Quantitative Structure–Activity Relationship models. *TrAC Trends Anal Chem.* 2013;42:49–63.

- [78] Shahlai M. Descriptor selection methods in quantitative structure-activity relationship studies: a review study. *Chem Rev.* 2013;113(10):8093-103.
- [79] Sutter JM, Kalivas JH. Comparison of forward selection, backward elimination, and generalized simulated annealing for variable selection. *Microchem J.* 1993;47(1-2):60-6.
- [80] Draper NR, Smith H, Pownell E. *Applied Regression Analysis.* New York: Wiley; 1966.
- [81] Weisberg S. *Applied Linear Regression.* New Jersey: John Wiley & Sons; 2005.
- [82] O'Brien SE, de Groot MJ. Greater than the sum of its parts: combining models for useful ADMET prediction. *J Med Chem.* 2005;48(4):1287-91.
- [83] Yap CW, Li H, Ji ZL, Chen YZ. Regression methods for developing QSAR and QSPR models to predict compounds of specific pharmacodynamic, pharmacokinetic and toxicological properties. *Mini Rev Med Chem.* 2007;7(11):1097-107.
- [84] Rucker C, Rucker G, Meringer M.  $\gamma$ -Randomization and its variants in QSPR/QSAR. *J Chem Inf Model.* 2007;47(6):2345-57.
- [85] Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb Sci.* 2003;22(1):69-77.
- [86] Hawkins DM, Basak SC, Mills D. Assessing model fit by cross-validation. *J Chem Inf Comput Sci.* 2003;43(2):579-86.
- [87] Koenig J. Does process excellence handcuff drug development? *Drug Discov Today.* 2011;16(9-10):377-81.
- [88] van de Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov.* 2003;2(3):192-204.
- [89] Lombardo F, Gifford E, Shalaeva MY. In silico ADME prediction: data, models, facts and myths. *Mini Rev Med Chem.* 2003;3(8):861-75.
- [90] United States Government Accountability Office (GAO). Report to Congressional Requesters. New Drug Development. Science, business, regulatory, and intellectual property issues cited as hampering drug development efforts. 2006.

- [91] Filmore D, Thayer AM, Willis RC. Pipeline challenges. Major pharmaceutical and biotechnology companies take a variety of approaches to remain productive. *Modern Drug Discov.* 2004;7:28-34.
- [92] Pammolli F, Magazzini L, Riccaboni M. The productivity crisis in pharmaceutical R&D. *Nat Rev Drug Discov.* 2011;10(6):428-38.
- [93] Woosley RL, Cossman J. Drug development and the FDA's Critical Path Initiative. *Clin Pharmacol Ther.* 2007;81(1):129-33.
- [94] Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov.* 2004;3(8):711-5.
- [95] Food and Drug Administration. Challenge and opportunity on the critical path to new medical products. In: Services UDoHaH, editor. 2004.
- [96] European Commission, European Federation of Pharmaceutical Industries and Associations. The Innovative Medicines Initiative (IMI) Strategic Search Agenda. 2006.
- [97] Bharath EN, Manujula SN, Vijaychand A. In silico drug design tool for overcoming the innovation deficit in the drug discovery process. *Int J Pharm Pharmaceut Sci.* 2011;3(2):8-12.
- [98] Oprea TI. Virtual Screening in Lead Discovery: A Viewpoint. *Molecules* 2002;7(1):51-62.
- [99] Talevi A, Bellera CL, Di Ianni M, Duchowicz PR, Bruno-Blanch LE, Castro EA. An integrated drug development approach applying topological descriptors. *Curr Comput Aided Drug Des.* 2012;8(3):172-81.
- [100] Salum LB, Andricopulo AD. Fragment-based QSAR: perspectives in drug design. *Mol Divers.* 2009;13(3):277-85.



## Capítulo IV

# Materiales y métodos

---



## 4.1 Conjunto de datos

En primer lugar se construyó el set o conjunto de datos, que consiste en un conjunto de ligandos los cuales, según datos experimentales, presentan y no presentan la actividad o propiedad de interés. A tal fin se realizó una búsqueda bibliográfica y a partir de la información obtenida de un total de 154 publicaciones, el conjunto de datos inicial quedó integrado por 305 sustratos y no sustratos de la BCRP con sus respectivas estructuras moleculares y datos de interacción con la proteína. En este punto, se consideró a un compuesto como sustrato sólo si es transportado por la BCRP y como un no sustrato en el caso contrario. Los compuestos que se unen a la BCRP pero no son transportados se consideraron correspondientemente como no sustratos. Debido tanto a la variedad de condiciones experimentales utilizadas para las determinaciones de las tasas de eflujo encontradas en la literatura como a la gran variabilidad inter-laboratorio reportada para los estudios experimentales de transporte por bombas MDR-ABC [1-2] no fue posible establecer un único valor de corte de referencia para identificar sustratos y no sustratos; en cambio, se tuvo en cuenta el criterio establecido en cada estudio original.

A partir del análisis exhaustivo de las 154 publicaciones compiladas surgió la necesidad de establecer un criterio de inclusión y exclusión adicional para definir cuáles de los compuestos del conjunto de datos original se utilizarán para desarrollar los modelos computacionales. Para ello se tomaron en consideración particularmente los siguientes factores:

- 1.- Cambios en la especificidad de sustrato dependientes de variantes genéticas del blanco molecular: una sustitución de nucleótido único en el aminoácido 482 (R482) causa cambios conformacionales que modifican notablemente la afinidad y especificidad de la BCRP por algunos sustratos e incluso su capacidad de eflujo [3-15]. Si bien existen otras mutaciones [11-12, 16], la del aminoácido 482, denominada mutación de alta frecuencia (en inglés, *hotspot mutation*), se ha convertido en el caso de mayor estudio y relevancia dado que se asocia a los mayores cambios en la especificidad de sustrato y capacidad de eflujo detectados hasta el momento, llegando a ampliar la especificidad de sustrato del transportador, es decir, el mutante es capaz de transportar compuestos no

transportados por la forma salvaje de la proteína [3-5, 9, 13-14, 17]. La BCRP de tipo salvaje presenta una arginina en la posición 482, mientras que la BCRP mutada en esa posición puede presentar una glicina o treonina en el caso de la proteína humana, o bien una serina o metionina en el caso de la *bcrp1* de ratón [18]. Los estudios de transporte *in vitro* de la BCRP mutada en la posición 482 indican que esta mutación otorga una ganancia en la función de la proteína tanto por ampliar el espectro de sustratos que es capaz de reconocer como por aumentar la velocidad de transporte y actividad ATPasa [6-9, 12]. Dicha mutación es altamente frecuente en cultivos celulares *in vitro* sometidos a altas concentraciones de fármacos citotóxicos y aún no se ha podido identificar su presencia *in vivo* (ni en la población sana ni en pacientes con distintos tipos de cáncer) [19-26] por lo que las posibles consecuencias clínicas de esta mutación aún no están claras y podría tratarse simplemente de un artefacto que ocurre como respuesta adaptativa al fuerte efecto mutagénico de los fármacos citotóxicos y/o a la pérdida del entorno biológico que experimentan los cultivos celulares *in vitro* [18].

2.- Cambios en la especificidad de sustrato dependientes del sistema de expresión: se han observado cambios en la especificidad de sustrato de la BCRP dependientes del sistema celular de expresión que se utilice en los ensayos de transporte *in vitro* para definir si un compuesto determinado es sustrato transportable, como pueden ser células de mamífero, de insectos o bacterias, entre las más comúnmente utilizadas. Esto puede deberse a diversos factores tales como cambios post-traduccionales dependientes del sistema de expresión utilizado [11, 27-40].

3.- Cambios en la especificidad de sustrato dependientes de la especie: si bien la *bcrp1* de rata y de ratón comparten un alto grado de identidad de la secuencia proteica de la BCRP humana (81% en ambos casos) y la especificidad de sustrato es muy similar entre las distintas especies [41], se han reportado algunas diferencias inter-especies [11, 27-39].

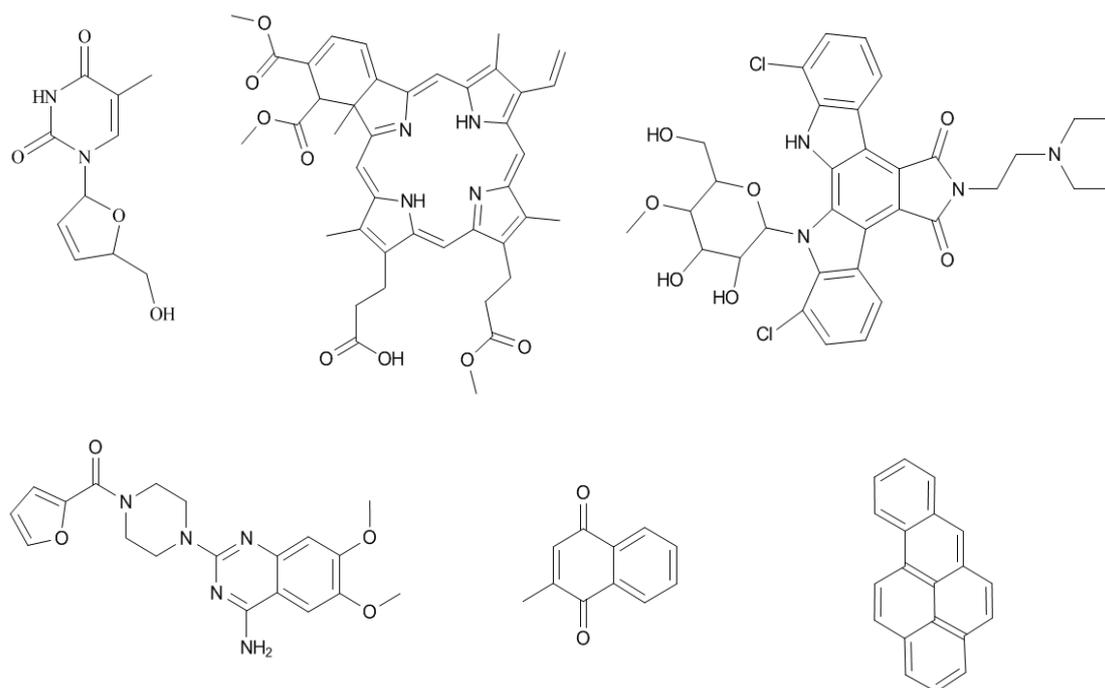
En función de estas variables y de su relevancia, se estableció como criterio adicional seleccionar y utilizar para la construcción del conjunto de datos final solamente los datos de aquellas publicaciones que utilizaron la BCRP humana de tipo salvaje expresada en líneas celulares humanas para la evaluación del

transporte de los compuestos. Sustratos y no sustratos de homólogos de otras especies de la BCRP sin evidencia de transporte mediado por la BCRP humana de tipo salvaje y de la BCRP humana expresada en células de otras especies, no se incluyeron para evitar el ruido asociado a la variabilidad de la especificidad de sustrato mencionada. Para seleccionar aquellos compuestos químicos que cumplan con el criterio establecido, se realizó una revisión exhaustiva y posterior clasificación de los datos extraídos de literatura y de esta manera a partir del conjunto de datos original de 305 compuestos químicos, se seleccionaron 262 compuestos, 156 sustratos y 106 no sustratos de la BCRP humana de tipo salvaje (Figura 4.1).



**Figura 4.1.** Esquema seguido para la construcción del set de datos final.

Es importante volver a destacar la amplia diversidad estructural de los compuestos que son reconocidos por la BCRP (ver ejemplos en la Figura 4.2). Esto mantiene una relación directa con la poliespecificidad que caracteriza a estos transportadores, ya discutida anteriormente, y a su vez permite asegurar un amplio dominio de aplicación de los modelos computacionales que se deriven de este conjunto de datos. En el Anexo I se presentan todas las estructuras químicas de los compuestos que constituyen el conjunto de datos final.



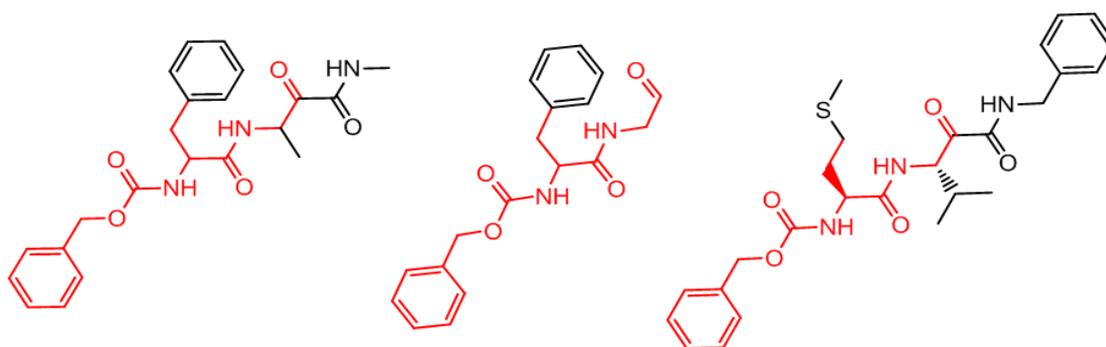
**Figura 4.2.** Estructuras químicas de 6 sustratos transportados por la BCRP donde se puede apreciar la diversidad estructural que es reconocida por el transportador.

## 4.2 Partición del conjunto de datos en los conjuntos de entrenamiento y de prueba

A partir del set de datos de 262 compuestos químicos se obtuvieron el conjunto de entrenamiento (en inglés, *training set*) a partir del cual se inferirá estadísticamente el modelo, y un conjunto de prueba (en inglés, *test set*) que se usará para estimar la capacidad predictiva del modelo, es decir la capacidad del modelo de predecir la categoría de casos no incluidos en el conjunto de entrenamiento.

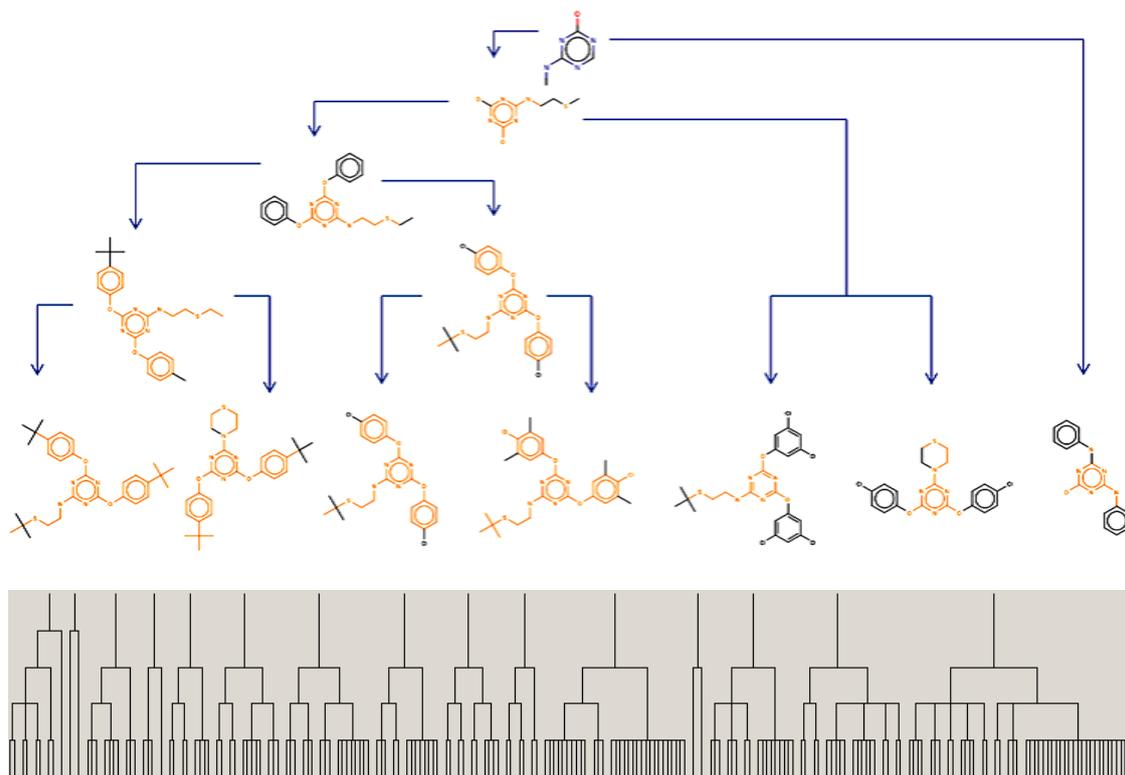
En esta instancia, un problema relevante es cómo particionar el conjunto de datos en los conjuntos de entrenamiento y de prueba de modo que éstos sean representativos del total de la información química contenida en el conjunto de datos. Se ha demostrado que la partición al azar es una aproximación adecuada cuando se seleccionan conjuntos de entrenamiento y de prueba de tamaño similar, lo que puede hacerse cuando se cuenta con un conjunto de datos de gran tamaño.

En cambio, cuando se trabaja con conjuntos de datos relativamente pequeños como en nuestro caso, donde se reserva la mayor proporción de los compuestos para el conjunto de entrenamiento a los fines de contar con la mayor información posible para entrenar los modelos y el conjunto de prueba suele ser más pequeño, aproximaciones de muestreo racionales proveen mejores resultados [42-43]. Siguiendo este último criterio, se utilizó la herramienta de agrupamiento (en inglés, *clustering*) jerárquico LibraryMCS v0.7 (ChemAxon) en combinación con el algoritmo de optimización de agrupamiento k-means del módulo Cluster Analysis del programa Statistica 10 (Statsoft Inc, 2011). La idea fundamental es identificar, dentro de una base de datos químicamente diversa, agrupamientos de estructuras con características químicas comunes a fin de guiar la selección de conjuntos de entrenamiento y de prueba representativos, garantizando así la máxima cobertura posible del espacio químico en uno y otro conjunto. LibraryMCS se basa en la máxima subestructura común (en inglés, *maximum common substructure*, MCS) obtenida a partir de pares de estructuras similares para agrupar conjuntos de estructuras sin recurrir a una comparación exhaustiva (y por ende computacionalmente demandante). Una subestructura común es definida como una subestructura presente en dos moléculas, conteniendo los mismos tipos de átomos y de enlaces químicos. La MCS es la subestructura común que contiene el mayor número de átomos, es decir, la de mayor tamaño (ver como ejemplo Figura 4.3) [44].



**Figura 4.3.** Se presenta coloreada en rojo, en las tres estructuras, un ejemplo de MCS.

LibraryMCS construye una matriz de similitud para el conjunto de estructuras consideradas y en función de los coeficientes de similitud [45] elige el par de estructuras más similares, considerando que las dos estructuras con el mayor coeficiente de similitud tienen mayor probabilidad de presentar la MCS. Una vez que esta probable MCS se ha establecido, todos los compuestos del conjunto de datos que la incluyan se agrupan conjuntamente. El proceso se repite iterativamente hasta que no se encuentran más pares de estructuras con un valor del coeficiente de similitud por encima del umbral utilizado por el algoritmo o hasta que quedan estructuras sin agrupar que no presentan una MCS de igual o mayor tamaño a la especificada por el usuario, en cuyo caso tales estructuras se presentan aisladas y constituyen compuestos atípicos (en inglés, *outliers*), es decir, no se incluyen en ninguno de los agrupamientos (en inglés, *clusters*). La búsqueda de MCS basada en similitud química conduce a soluciones reproducibles pero aproximadas, ya que no hay garantía de que aquellas estructuras con mayor valor del coeficiente de similitud contengan, efectivamente, la MCS [46]. En este caso, se fijó que la MCS en torno a la cual se desarrolló el agrupamiento jerárquico debía tener al menos 9 átomos (valor por default del programa utilizado). La Figura 4.4 presenta un ejemplo de agrupamiento a través de esta aproximación.



**Figura 4.4.** Ejemplo de agrupamiento basado en MCS. Las estructuras del conjunto de datos que no poseen una subestructura común con el resto con un mínimo número de átomos especificado por el usuario, son representadas de manera aislada en el dendrograma.

Seguidamente los grupos de compuestos así obtenidos fueron optimizados utilizando el algoritmo k-means. El método k-means implica un algoritmo de optimización de agrupamientos que –como cualquier algoritmo de optimización de agrupamientos– obedece los siguientes pasos generales: a) encontrar o proponer una partición inicial de  $n$  objetos en  $g$  grupos; b) calcular la variación en un criterio de agrupamiento cuando un objeto de un grupo se mueve a otro; c) realizar el cambio que conduce a la mayor mejora en el valor del criterio de agrupamiento; d) repetir el proceso anterior hasta que ningún desplazamiento de objetos de un grupo a otro produzca una mejoría. El método k-means inicia con un número de grupos definidos por el usuario y mueve los objetos entre esos grupos con el objetivo específico de 1) minimizar la variabilidad dentro de los agrupamientos y 2) maximizar la variabilidad entre los agrupamientos. Para esto utiliza la distancia euclidiana al cuadrado para la medida de las distancias entre los objetos y el

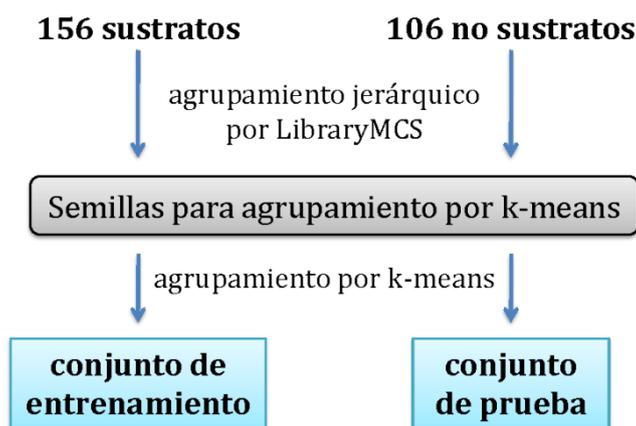
centroide de los agrupamientos [47-48]; el criterio de agrupamiento es la minimización de la suma de cuadrados intragrupo [49]:

$$\sum_{m=1}^g \sum_{l=1}^{n_m} (x_{ml} - \bar{x}_m)^2$$

donde  $g$ , como se dijo, representa el número de grupos en los que se particionan los objetos,  $n_m$  es el número de objetos en el grupo  $m$ ,  $x_{ml}$  denota el vector  $p$ -dimensional de observaciones del objeto  $l$  del grupo  $m$ , y  $\bar{x}_m$  es el vector que representa la media de los elementos del grupo  $m$  en el espacio  $p$ -dimensional (considerando que se utilizan  $p$  variables, por ejemplo  $p$  descriptores moleculares, para caracterizar cada objeto). A partir de la partición inicial, cada objeto se realocaliza hacia el grupo cuya media le resulta más próxima, y el proceso se repite hasta encontrar convergencia. En este caso, el número de grupos de la partición inicial se definió a partir del resultado del agrupamiento jerárquico de LibraryMCS. Para esto, las “semillas” (elementos que se asignan inicialmente a los grupos y a partir de los cuales se construyen los agrupamientos iniciales) fueron elegidas al azar a partir de los agrupamientos obtenidos vía LibraryMCS para las categorías sustratos y no sustratos. En el programa Statistica se utilizaron 10 iteraciones hasta convergencia. Para caracterizar a cada objeto (compuesto químico) y definir el espacio  $p$ -dimensional en el que se aplicaría  $k$ -means, se seleccionaron distintos descriptores moleculares representativos de diferentes aspectos de la molécula calculados con Dragon 4.0 (Milano Chemometrics, 2003): peso molecular, logaritmo del coeficiente de reparto octanol-agua de Moriguchi ( $mlogP$ ), área de la superficie polar, número de grupos aceptores y donantes de enlaces de hidrógeno, índice de información del contenido atómico, suma de volúmenes atómicos de van der Waals, suma de electronegatividades atómicas de Sanderson, y el índice de forma 2D de Petitjean. Estos descriptores fueron normalizados y aplicados para el cálculo de la distancia de cada elemento a la media de su grupo tentativo. De esta manera se obtuvieron separadamente los agrupamientos de las categorías sustrato y no sustrato a partir de los cuales se obtuvieron los conjuntos de entrenamiento y de prueba. Se destinó el 50% de cada agrupamiento de los sustratos y el 75% de

cada agrupamiento de los no sustratos al conjunto de entrenamiento, mientras que para el conjunto de prueba esos porcentajes fueron del 50% para los sustratos y del 25% para los no sustratos. De este modo se obtuvo un conjunto de prueba balanceado.

La Figura 4.5 muestra una representación esquemática de la metodología de agrupamiento utilizada para la partición racional del conjunto de datos.



**Figura 4.5.** Esquema de la metodología de agrupamiento aplicada para la partición del conjunto de datos en el conjunto de entrenamiento y el conjunto de prueba.

### 4.3 Cálculo de los descriptores moleculares

Para el cálculo de los descriptores moleculares se utilizaron dos versiones del programa Dragon. Por un lado se utilizó la versión 4.0 (Milano Chemometrics, 2003) el cual permite el cálculo de más de 1.600 descriptores moleculares (0-3D) que están divididos en 20 bloques lógicos. De los más de 1600 descriptores que posee el programa se excluyeron los descriptores tridimensionales y se calcularon para cada compuesto 867 descriptores de baja dimensionalidad más el bloque de propiedades moleculares, los cuales se encuentran distribuidos a lo largo de 13 de los 20 bloques de descriptores que presenta Dragon 4.0.

Por otro lado se utilizó la versión 6.0 (Milano Chemometrics, 2012) la cual, al ser una versión posterior del software, permite el cálculo de un mayor número de

descriptores moleculares: más de 4800 (0-3D) que en este caso están divididos en 29 bloques lógicos. Nuevamente los descriptores tridimensionales fueron excluidos del cálculo y se calcularon 2029 descriptores, los cuales se encuentran distribuidos a lo largo de 19 de los 29 bloques de descriptores que presenta esta versión de Dragon.

En este punto es muy importante eliminar los descriptores de los compuestos del conjunto de entrenamiento que tienen baja variabilidad y por lo tanto contienen poca o ninguna información y carecen de utilidad. Para esto se aplicaron los filtros iniciales proporcionados por el software Dragon de manera de excluir los descriptores moleculares con valores constantes o casi constantes (valores idénticos para todos compuestos del conjunto de entrenamiento menos uno) y los descriptores con desviación estándar por debajo de 0.001.

#### ***4.4 Método de modelado***

Como se anticipó en el capítulo anterior, la presente tesis está enfocada en el desarrollo de modelos clasificadores los cuales constituyen modelos cuantitativos basados en relaciones entre las variables independientes (en este caso los descriptores moleculares) y una variable dependiente de respuesta categórica (en este caso binaria) que puede ser nominal (aquí, "sustrato" y "no sustrato") o valores numéricos enteros (aquí, "1" y "-1") que representan la categoría de clase del objeto que está siendo clasificado. El término "cuantitativo" se refiere al valor numérico de las variables independientes [50], de modo que los clasificadores proporcionan una respuesta cualitativa del tipo si/no (es decir, sustrato o no sustrato) al tiempo que conservan el análisis cuantitativo a través de las variables independientes numéricas [51]. En comparación con los modelos de variable dependiente continua, este tipo de modelos pueden ser útiles para manejar datos ruidosos, como es este caso en donde se compilieron datos experimentales de la bibliografía sujetos a gran variabilidad inter-laboratorio. Los modelos de variable dependiente continua presentan como salida o respuesta un valor (por ejemplo de IC50) que se espera sea una explicación o predicción exacta. Por este motivo, el rendimiento de un modelo con una salida continua depende fuertemente de la

calidad de los datos experimentales de los que fue derivado. Sin embargo, el rendimiento de los modelos clasificadores está menos influenciado por la variabilidad experimental ya que las fuentes de ruido están particularmente restringidas a aquellos puntos que se encuentran en la frontera entre las categorías de objetos en cuestión (por ejemplo, aquellos compuestos cuyo nivel de transporte se encuentra cerca del umbral que se ha definido para diferenciar sustratos de no sustratos) [52].

Uno de los principales desafíos a enfrentar a la hora de encarar un proceso de modelado es definir qué tipo de modelo matemático (en este caso, algoritmo de clasificación) es el más adecuado para el tema que se pretende modelar, de modo que el modelo obtenido permita ajustar adecuadamente los datos, explicando el comportamiento de los compuestos del conjunto de entrenamiento de forma eficaz y prediciendo datos externos de una forma lo suficientemente confiable. Como ya se mencionó anteriormente, ninguna técnica es consistentemente mejor que todas las demás y las comparaciones sistemáticas de diferentes técnicas de modelado con diferentes niveles de sofisticación han demostrado sorprendentemente que los enfoques menos sofisticados a veces pueden superar los más complejos en términos de las métricas de enriquecimiento, esto es, los métodos más complejos no son inequívocamente mejores y qué método es el más adecuado parece ser altamente dependiente del sistema bajo estudio, mientras que por otro lado los enfoques más simples son siempre más eficientes [53-56].

Siguiendo esta línea de pensamiento, en el presente trabajo se encararon dos instancias diferentes de modelado; en la primera instancia se utilizaron métodos más sencillos y luego, en función de los resultados obtenidos, se decidió avanzar a una segunda instancia donde se abordaron métodos más complejos, quedando definidas de la siguiente manera:

- 1) Primera instancia: construcción de modelos lineales
- 2) Segunda instancia: construcción de modelos no lineales.

### 4.4.1 Primera instancia de modelado: construcción de modelos lineales

#### 4.4.1.1 Primera serie de modelos lineales

Los primeros modelos lineales se construyeron utilizando análisis lineal discriminante (ALD) con lo cual se obtuvieron y caracterizaron las correspondientes funciones discriminantes lineales ( $Df$ ). El ALD es una metodología de aprendizaje supervisado sencilla que consiste en derivar, mediante el método de los cuadrados mínimos, una combinación lineal ( $Df$ ) de variables independientes (los descriptores moleculares) capaces de diferenciar entre objetos de dos o más categorías o clases (aquí, sustratos y no sustratos de la BCRP) [52]. Adopta la siguiente forma:

$$Df = a_0 + \sum_i a_i d_i$$

donde  $a_0$  es una constante y  $a_i$  es el coeficiente asociado con el descriptor molecular  $d_i$ . La función obtenida  $Df$  corresponde a un plano en el espacio  $k$ -dimensional (siendo  $k$  el número de descriptores incluidos en el modelo) que, idealmente, deja a uno y otro lado los compuestos sustratos y no sustratos. Como se ve, el ALD es muy similar a la regresión lineal múltiple excepto por el hecho que la variable dependiente observada es ahora una variable categórica que sirve para etiquetar las clases y puede asumir tantos valores como categorías se estén considerando. En este caso, como se consideran dos categorías la variable dependiente es de naturaleza binaria, asumiendo de manera arbitraria, un valor de 1 para los compuestos de la categoría sustratos y un valor de -1 para los compuestos de la categoría no sustratos. Debido a los valores arbitrarios asignados a la variable dependiente, los sustratos tenderán a tener valores positivos de  $Df$ , y los no sustratos tenderán a adoptar valores negativos.

Como ya se analizó en el Capítulo III, una etapa fundamental del modelado es la adecuada selección de variables, es decir, la selección -a partir del total de

descriptores moleculares disponibles- del subconjunto de descriptores a incluir en el modelo que mejor se correlacionan con la propiedad de interés. Dado que cuando se trabaja con un número alto de descriptores (como en nuestro caso 867 calculados con Dragon 4.0 y 2029 con Dragon 6.0) pueden aparecer correlaciones azarosas entre la propiedad modelada y un subconjunto de descriptores, se utilizó en esta primera instancia de modelado la estrategia denominada, por algunos autores, como *random subspace* [52, 57], que consiste en explorar subconjuntos aleatorios más pequeños de descriptores moleculares del conjunto total de descriptores. Para ello se generaron 102 subconjuntos de descriptores a partir de combinaciones aleatorias de los bloques de descriptores de Dragon4.0, donde cada subconjunto contiene no más de 200 descriptores moleculares. Sólo a modo de ejemplo, el primer subconjunto (de 180 descriptores) surgió de la combinación de los siguientes bloques: descriptores constitucionales, índices basados en autovalores, autocorrelaciones 2D y propiedades moleculares; el segundo subconjunto (de 195 descriptores) combinó a los bloques: recuento de trayectos y caminos, índices de conectividad, recuento de los grupos funcionales y descriptores BCUT; y así sucesivamente.

Para seleccionar los descriptores de cada subconjunto que mejor discriminen la categoría de los compuestos, se debe ponderar la incidencia de cada descriptor en la variable dependiente. La evaluación exhaustiva de todas las combinaciones posibles de descriptores exigiría obtener y evaluar  $D!/[d!(D-d)!]$  combinaciones, donde  $D$  es el número de descriptores en un subconjunto de descriptores dado y  $d$  es el número de descriptores incluidos en un modelo dado. Esto es computacionalmente muy exigente o incluso inviable cuando  $D$  es grande, como es nuestro caso. Por lo tanto, para la construcción de la  $Df$  se utilizó la técnica por pasos hacia adelante (en inglés, *forward stepwise*), un enfoque por etapas que, aunque más rápido, conduce habitualmente a soluciones subóptimas [52]. Mediante esta técnica la  $Df$  se construye paso a paso y las variables (descriptores) se introducen en el modelo de a una a la vez. Por lo tanto, el modelo comienza sin descriptores y en cada paso se revisan todos los descriptores y se evalúa cuál es el que más contribuye a la discriminación entre los grupos. Ese descriptor se incluye en el modelo y se procede a la siguiente etapa. Una vez que un descriptor es incorporado al modelo no puede ser eliminado en los pasos

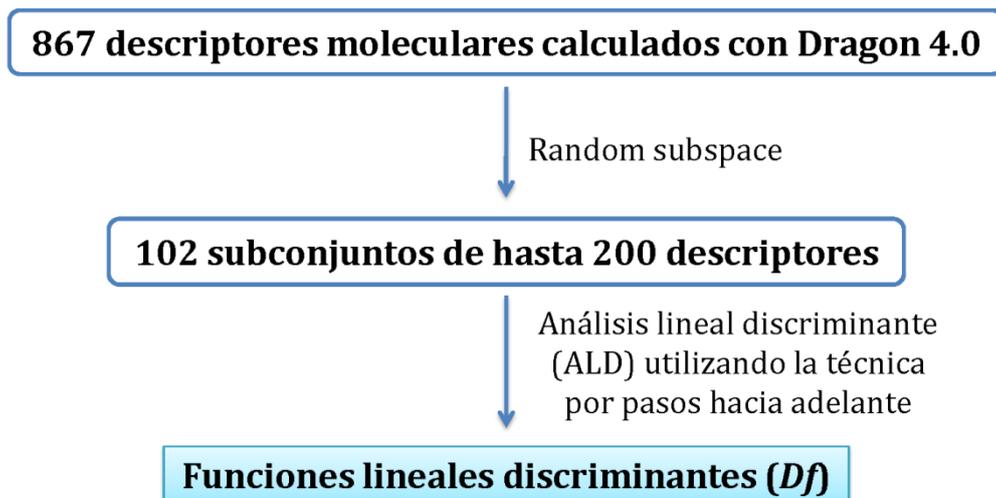
subsiguientes. El procedimiento por pasos está "guiado" por los respectivos valores de  $F$  para entrar y  $F$  para eliminar. El valor  $F$  para una variable indica su significancia estadística en la discriminación entre grupos, es decir, es una medida de la magnitud en que un descriptor hace una contribución única a la predicción de pertenencia al grupo. En general, se continúa incorporando descriptores al modelo, siempre que los respectivos valores de  $F$  para esas variables sean más grandes que el  $F$  para entrar especificado por el usuario (que en este caso fue de 1). Además se exigió una relación de casos/descriptores de 10 o más, es decir al menos 10 compuestos en el conjunto de entrenamiento por cada descriptor incorporado al modelo. De este modo, el máximo número de descriptores que se admitió incorporar en un modelo fue de 15, reduciendo así la probabilidad de sobreajuste del modelo (ganancia en la capacidad explicativa del modelo a expensas de su capacidad predictiva) [58-60].

Para obtener los modelos se utilizó el módulo de análisis lineal discriminante del programa Statistica 10. Se utilizaron valores de tolerancia no inferiores a 0.1 con el fin de limitar la posibilidad de inclusión de pares de descriptores altamente correlacionados o redundantes. La tolerancia  $\beta$  se define como:

$$\beta_j = 1 - R_j^2$$

Donde  $R_j^2$  es el mayor coeficiente de determinación entre una variable independiente del modelo y todas las demás. Finalmente, se consideraron sólo aquellos modelos que incluyeron descriptores con coeficientes significativos a un nivel  $\alpha = 0.05$ .

En la Figura 4.6 se presenta un esquema del procedimiento de modelado seguido en esta etapa.



**Figura 4.6.** Representación esquemática simplificada de los pasos seguidos para la obtención de la primera serie de modelos lineales.

#### 4.4.1.2 Segunda serie de modelos lineales

En esta oportunidad se utilizó el método de sustitución mejorado o ERM (por el inglés, *Enhanced Replacement Method*) [50, 61-62] para seleccionar, de entre los 2029 descriptores calculados por Dragón 6.0, las combinaciones lineales de descriptores capaces de predecir la categoría de los compuestos. El método de sustitución (*Replacement Method*, RM) original fue desarrollado para explorar el espacio de los descriptores de una manera eficiente, en busca de un subconjunto de descriptores moleculares a partir de un gran conjunto de descriptores [52, 61, 63]. Se trata de un algoritmo iterativo que converge rápidamente y que produce modelos lineales con un desempeño muy cercano al obtenido con una búsqueda exhaustiva pero con mucho menos costo computacional y, según los autores que lo desarrollaron, mejor que el de los más elaborados algoritmos genéticos [50]. Básicamente se selecciona al azar un subconjunto inicial de  $d$  descriptores del conjunto de descriptores totales  $D$ . A continuación, uno de los descriptores seleccionados,  $X_i$ , se sustituye uno a uno con todos los  $D - d$  descriptores

restantes del conjunto de descriptores totales. El subconjunto de descriptores que presenta menor desviación estándar ( $S$ ) se conserva. De este subconjunto resultante, el descriptor con la mayor desviación en su coeficiente de regresión es nuevamente sustituido uno por uno con todos los  $D - d$  descriptores restantes. Este procedimiento se repite hasta que no se producen nuevas modificaciones en el subconjunto de descriptores seleccionados. En cada ciclo, los descriptores reemplazados en los pasos anteriores no se vuelven a considerar. Si la sustitución del descriptor que tiene el coeficiente con mayor desviación por aquellos descriptores del conjunto  $D - d$  no disminuye el valor de  $S$ , entonces ese descriptor no se sustituye, es decir, se retiene en el subconjunto final. Una versión mejorada del RM, denominada método de sustitución modificado (*Modified Replacement Method*, MRM [62]), sigue la misma estrategia excepto que en cada paso el descriptor con el mayor error es sustituido incluso si la sustitución no se ve acompañada por un menor valor de  $S$ . Esta modificación hace que el MRM presente menos probabilidades de quedar atrapado en un mínimo local a costa de una convergencia más lenta (mayor costo computacional). El ERM surge de la combinación del RM y el MRM siguiendo la secuencia RM-MRM-RM; que combina las buenas características de ambos métodos, desarrollando un ciclo de recocido simulado (en inglés, *simulated annealing*) completo [62]. ERM es menos dependiente del subconjunto de descriptores seleccionado al azar inicialmente y presenta menos tendencia a caer en mínimos locales.

Se construyeron modelos conteniendo entre 4 a 12 descriptores moleculares. No se consideraron modelos con un mayor número de descriptores de modo de reducir la posibilidad de sobreajuste, y al igual que en la etapa anterior sólo se consideraron aquellos modelos que incluyeron descriptores con coeficientes significativos a un nivel  $\alpha = 0.05$ .

#### *4.4.2 Segunda instancia de modelado: construcción de modelos no lineales*

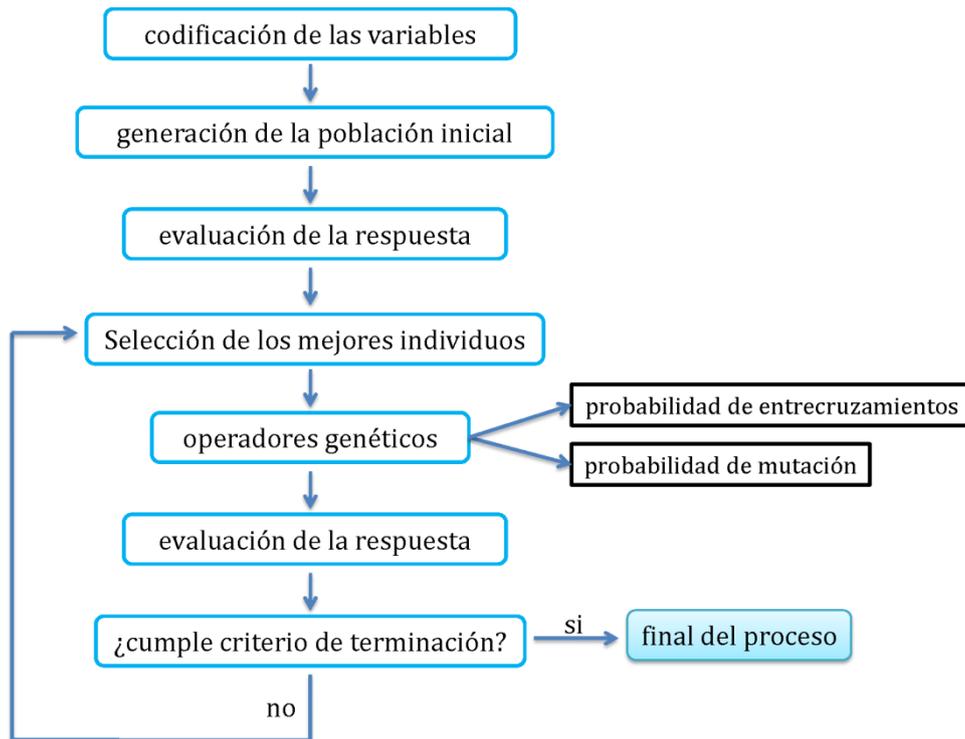
En esta instancia de modelado se construyeron árboles de decisión que son ampliamente conocidos por su capacidad predictiva, su facilidad de interpretación y su robustez para trabajar con datos ruidosos (como en nuestro caso) [64]. Para ello, en primer lugar se procedió a la aleatorización de los 867 descriptores totales calculados por Dragón 4.0 y posterior partición en 6 conjuntos de alrededor de 145 descriptores cada uno. La aleatorización permite la evaluación de conjuntos de descriptores de diversos bloques simultáneamente, en contraste a la metodología utilizada en el caso de la primera serie de modelos lineales, donde se respetaron los bloques lógicos de descriptores. La partición en conjuntos de menor tamaño nuevamente se debe a la necesidad de reducir el número de descriptores a ser analizados simultáneamente. La variable dependiente asignada en este caso fue una variable binaria nominal que representa las etiquetas de clases: "sustratos" y "no sustratos".

Se aplicaron algoritmos genéticos (GA) implementados en el software Weka 3.6 [65] para pre-seleccionar los descriptores de cada conjunto con mejor capacidad discriminante. De esa manera se redujo aún más el número de descriptores a explorar mediante el algoritmo J48 (implementación del algoritmo de inducción de árboles de decisión C4.5 [66] en Weka 3.6) para la obtención de los correspondientes árboles de decisión.

Los GA fueron propuestos por Holland en 1975 [67], pero recién fue posible aplicarlos en tiempos computacionales razonables desde la década de 1990, cuando las computadoras se hicieron más veloces [68-69]. Básicamente consisten en una técnica de optimización mediante la cual se trata de encontrar una solución a través de la búsqueda en el espacio de soluciones utilizando un algoritmo basado en la selección natural de la evolución biológica, es decir, en la teoría darwiniana de la "supervivencia del más apto". La evolución de las especies puede ser considerada por sí misma una forma de optimización, en donde la respuesta a ser optimizada es la aptitud para sobrevivir en determinado medio ambiente. La teoría de la evolución plantea que los individuos con una mayor "aptitud para el medio

ambiente" tienen una mayor probabilidad de sobrevivir y una mayor probabilidad de ganar las peleas por el apareamiento. De tal manera, el contenido genético de los mejores individuos estará cada vez más presente en las siguientes generaciones, ya que se transmite a través de la descendencia. Los GA resultan útiles para problemas complejos, como en este caso donde el número de variables independientes supera ampliamente el número de compuestos en el conjunto de entrenamiento [70]. Constan de cuatro pasos fundamentales: inicialización, selección, aplicación de los operadores genéticos y evaluación de la respuesta [71]. El algoritmo inicia codificando las potenciales variables independientes en secuencias binarias que asimilaremos alegóricamente a un "cromosoma". Cada cromosoma representa un modelo y cada "gen" del cromosoma se asocia a un descriptor molecular del pool de descriptores analizado. Si un determinado gen se encuentra asociado a un valor 0, se interpreta que el descriptor molecular aludido no se encuentra presente en el modelo correspondiente a ese cromosoma; si en cambio el gen se asocia a un valor 1, el descriptor aludido se encontrará incluido en el modelo. Se genera de esta forma una población inicial de "cromosomas" que varía de acuerdo al problema (generalmente entre 20 y 100). Posteriormente, los cromosomas de esta generación inicial se aparean o entrecruzan: en la prole de dos cromosomas apareados aparecerán mezcladas las características de sus progenitores. La selección se ocupa de la transferencia de una proporción de la población a la siguiente generación en cada generación sucesiva. Este proceso se realiza de acuerdo a una función de aptitud que permite la evaluación de la respuesta y es siempre dependiente del problema en estudio. La evaluación de la respuesta para cada cromosoma de cada generación permite seleccionar los mejores individuos y mediante la aplicación de los operadores genéticos se obtiene una nueva generación (soluciones hijas) a partir de los "padres" seleccionados. El operador genético más importante es por lo tanto la probabilidad de cruce o entrecruzamiento que opera sobre dos cromosomas a la vez para generar dos descendientes donde se combinan las características de ambos cromosomas padres, mientras que otro operador sumamente útil es la probabilidad de mutación que modifica al azar una parte de un cromosoma y permite alcanzar zonas del espacio de búsqueda que no estaban cubiertas por los individuos de la población original. Este proceso se repite un número determinado de

generaciones, ya sea hasta convergencia (donde las nuevas generaciones no resultan más aptas), hasta alcanzar un número máximo de generaciones establecido por el usuario o simplemente de acuerdo al tiempo total de corrida permitido [55]. La Figura 4.7 presenta un esquema de los pasos básicos que sigue un GA.



**Figura 4.7.** Esquema de las etapas básicas en las que consiste un GA.

Los parámetros utilizados en este caso fueron:

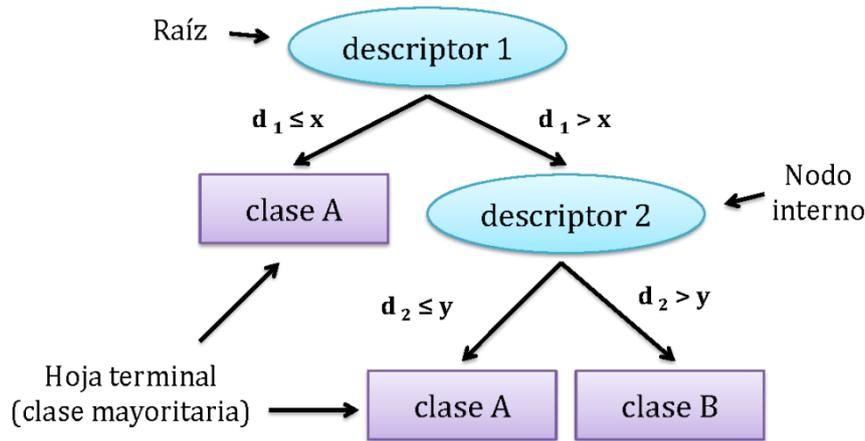
- Tamaño de la población inicial: 100
- Máximo número de generaciones: 50
- Probabilidad de entrecruzamientos: 0.6
- Probabilidad de mutaciones: 0.033

Como función de evaluación de la respuesta se utilizó el evaluador de la capacidad clasificadora del subconjunto definido por cada cromosoma (en inglés,

---

*Classifier subset evaluator*) sobre el conjunto de entrenamiento. Utiliza un algoritmo de inducción para estimar el "mérito" del subconjunto de descriptores, es decir, para evaluar la correspondencia de los descriptores con la clase. Como algoritmo de inducción se utilizó el mismo algoritmo J48 que luego se aplicó para caracterizar los correspondientes árboles de decisión finales.

Los árboles de decisión permiten obtener árboles de clasificación (o de regresión) mediante un método jerárquico divisivo. Para construir un árbol se utilizan reglas de división del tipo "si  $d > x$  entonces el elemento  $o$  pertenece a la clase A y si  $d \leq x$  entonces pertenece a la clase B" (siendo  $d$  en este caso un descriptor y  $x$  el valor de corte para ese descriptor) para realizar una serie de divisiones binarias de los datos en subconjuntos (nodos internos), hasta llegar a un árbol maximal donde se reparten todas las observaciones en la hojas o nodos terminales (ver Figura 4.8). En cada nodo del árbol, el algoritmo elige el descriptor que más eficazmente divide el conjunto de entrenamiento en subconjuntos enriquecidos en una clase u otra. El criterio de clasificación utilizado en el proceso de búsqueda es la "entropía", introducida por Shannon [72] y la "ganancia de información" [73]. La medida de la entropía se aplica de forma jerárquica en cada nivel del árbol de decisión y la ganancia de información se utiliza para medir la reducción esperada de la entropía en el nivel inmediato inferior de la jerarquía. En cada nivel de la jerarquía del árbol de decisión, el atributo que proporciona la mayor reducción de la entropía, es decir, la mayor ganancia de información, se selecciona como un nodo del árbol [74]. De esta manera el algoritmo divide los datos recursivamente en subgrupos más pequeños enriquecidos en una determinada categoría. Para el caso de la clasificación, cuando ninguno de los descriptores proporciona una ganancia de información, se crea un nodo terminal (hoja) donde se elige la clase que está más representada mediante un voto mayoritario [75].



**Figura 4.8.** Árbol de decisión. "x" e "y" representan el valor de corte del descriptor de cada nodo.

A los efectos de evitar el sobreajuste sobre el conjunto de entrenamiento y en aquellos casos en los que se trabaja con numerosos compuestos atípicos en el conjunto de entrenamiento (como en este caso), se utiliza un algoritmo aglomerativo (algoritmo de poda o *pruning* en inglés) que remueve varias hojas y ramas del árbol, obteniéndose un árbol que tenga mejor poder de predicción como así también menor tamaño y complejidad. Básicamente hay 2 métodos de poda [66, 76]:

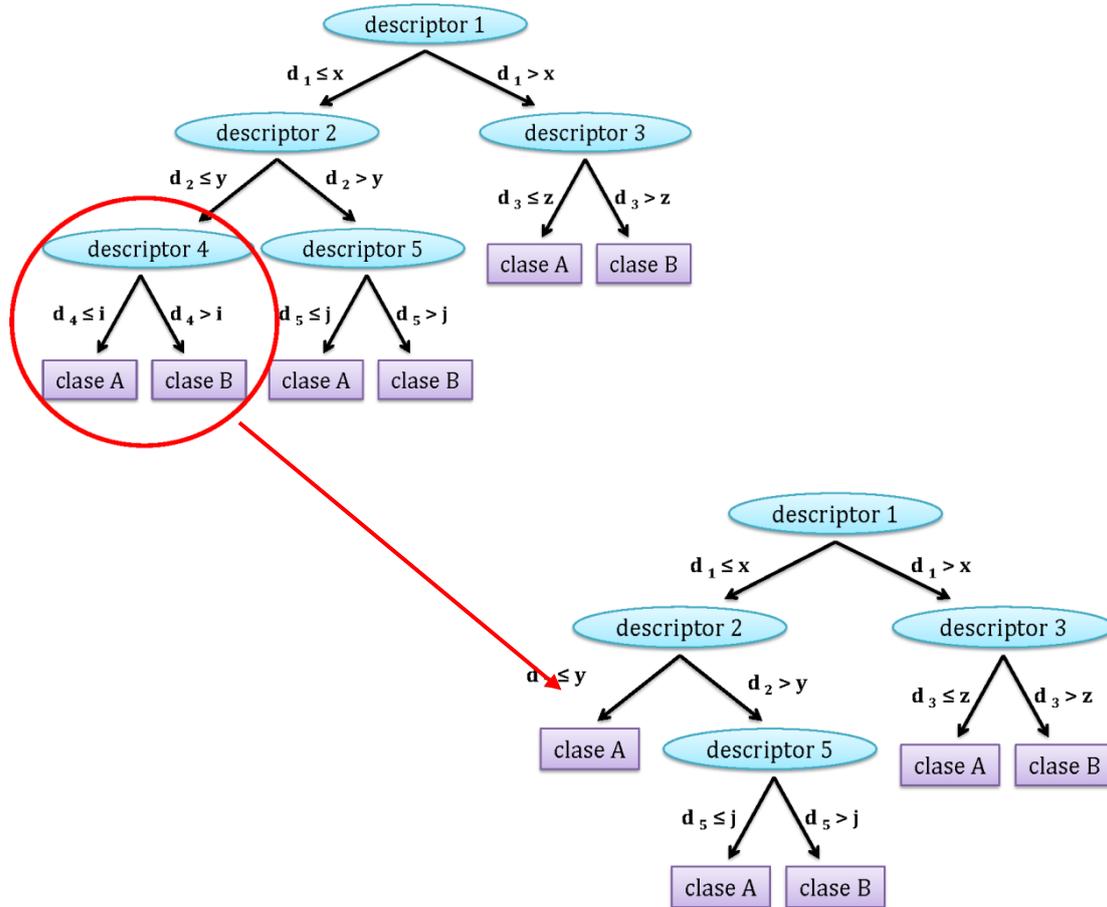
- post poda, que se lleva a cabo una vez que se finalizó la construcción del árbol
- poda en línea, que se realiza a medida que el árbol es inducido

El principio que gobierna al proceso de poda es comparar la cantidad de errores que un árbol de decisión comete antes y después de cada posible procedimiento de poda, de modo de reducir al máximo ese error. La métrica utilizada para describir la posibilidad de error (error estimado o  $E$ ), se calcula con la siguiente fórmula:

$$E = \frac{e + 1}{N + m}$$

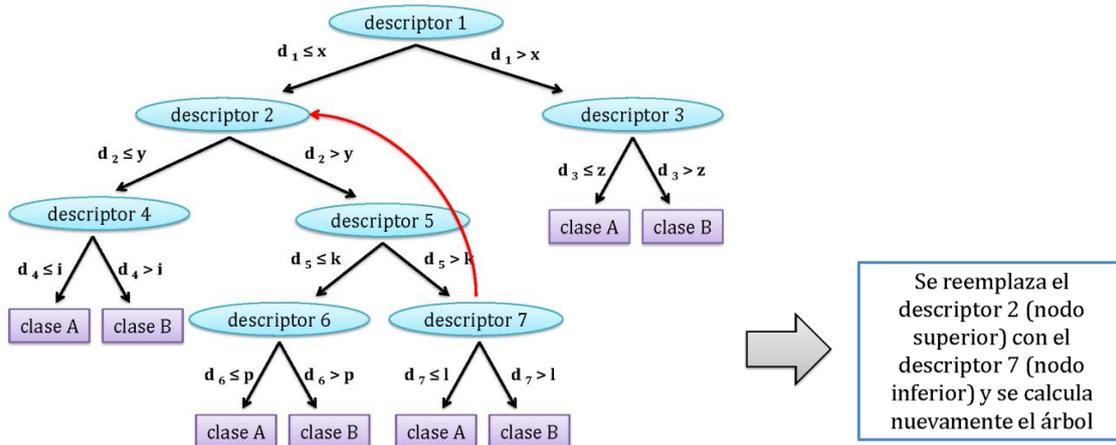
donde  $e$  son los compuestos mal clasificados en un nodo dado,  $N$  son los compuestos totales que llegan al nodo dado y,  $m$  son todos los compuestos del conjunto de entrenamiento. El valor tolerado para  $E$  se encuentra definido por el programa.

La post poda se realiza desde las hojas hacia la raíz del árbol y hay básicamente dos métodos: el reemplazo de subárboles (en inglés, *subtree replacement*) y la elevación de subárboles (en inglés, *subtree raising*). El reemplazo de subárboles (Figura 4.9) consiste en reemplazar nodos por hojas y depende del valor del factor de confianza (parámetro  $C$ ), donde menores valores de este parámetro (lo que implica en términos generales una menor confianza en los datos de entrenamiento) conlleva una mayor poda. En cada punto, el algoritmo compara el error promedio de cada nodo hijo versus el error si ese nodo hijo es eliminado y reemplazado por una hoja terminal con la etiqueta de clase de la clase mayoritaria. Los errores de clasificación de los datos de entrenamiento en cada nodo no proporcionan una estimación suficiente del error, dado que el árbol fue construido con estos datos, y no permitiría ninguna poda. En cambio, el error de clasificación debe ser entendido como una aproximación del error real basado en datos incompletos. Aquí es donde la noción de confianza estadística entra en juego. Se toma una suposición pesimista del error real de clasificación basado en los datos de entrenamiento. Si se tiene menos confianza en los datos de entrenamiento (que corresponde a un factor de confianza inferior), la estimación del error para cada nodo sube, aumentando la probabilidad de poda a favor de un nodo superior más estable. Los nodos con muy pocos compuestos son penalizados, ya que no se pueden hacer suposiciones seguras de su error de clasificación real.



**Figura 4.9.** Esquema donde se muestra el método de reemplazo de subárboles de post poda. "x", "y", "z", "i" y "j" representan el valor de corte del descriptor de cada nodo.

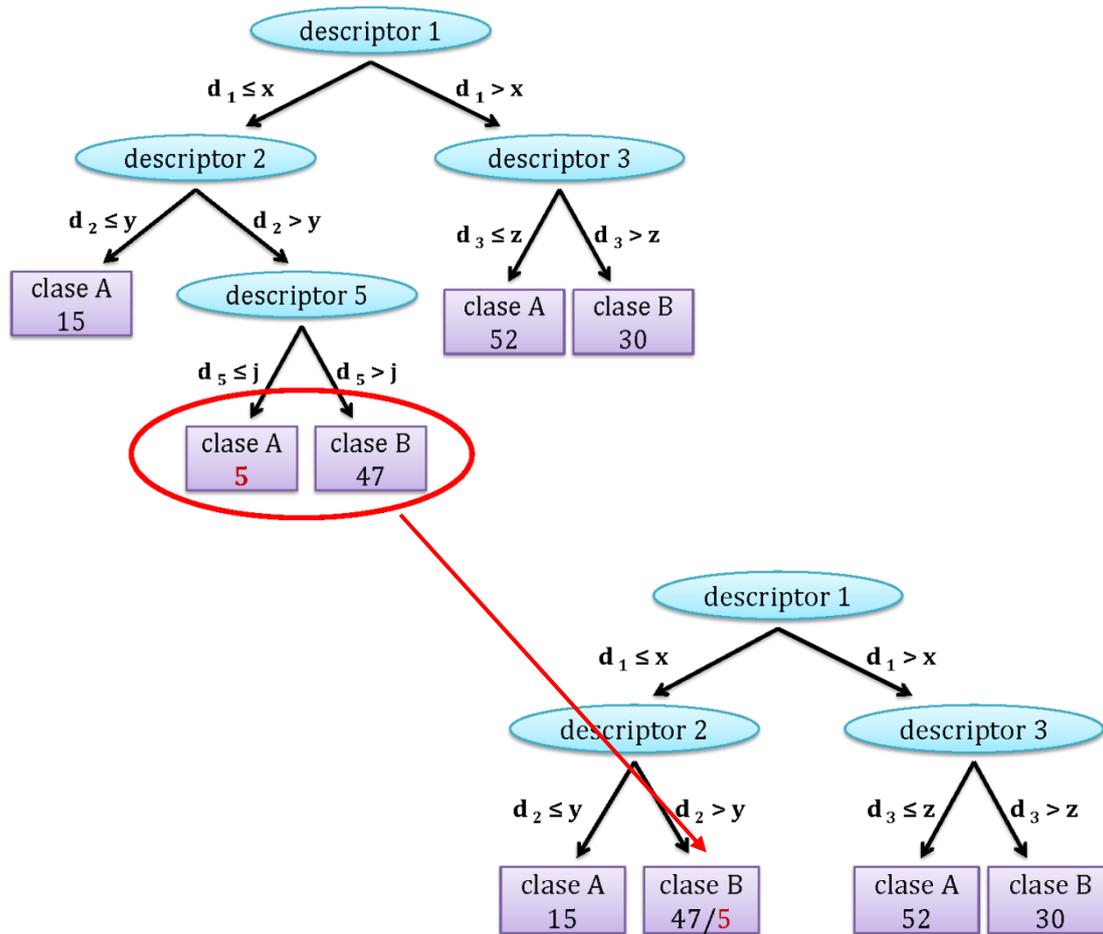
La elevación de subárboles (Figura 4.10) consiste en mover nodos que se encuentran más abajo en el árbol (es decir, más cerca de las hojas) hacia arriba (es decir, hacia la raíz) y de ese modo se calcula nuevamente el árbol reemplazando nodos superiores con nodos inferiores.



**Figura 4.10.** Esquema donde se muestra el método de elevación de subárboles de post poda. "x", "y", "z", "i", "k", "p" y "l" representan el valor de corte del descriptor de cada nodo.

Por otro lado, el método de poda de error reducido (en inglés, *reduced error pruning*, REP) es un método de post poda, donde se reserva una fracción del conjunto de entrenamiento (conjunto de poda o *pruning set*), se entrena el árbol con el conjunto de entrenamiento restante y finalmente se busca la versión más pequeña del árbol que mejor clasifica al conjunto de poda. Presenta la desventaja de trabajar con un conjunto de entrenamiento más reducido y por lo tanto menor información para la construcción del modelo.

La poda en línea (Figura 4.11) se realiza a medida que el árbol está siendo inducido, para lo que se establece un número mínimo de compuestos permitido por hoja (parámetro M). En cada proceso de división, cuando se excede este número se crea un nuevo nodo, mientras que cuando no es alcanzado, la división no se produce, los datos forman una hoja terminal y la etiqueta de clase corresponde a la clase mayoritaria.



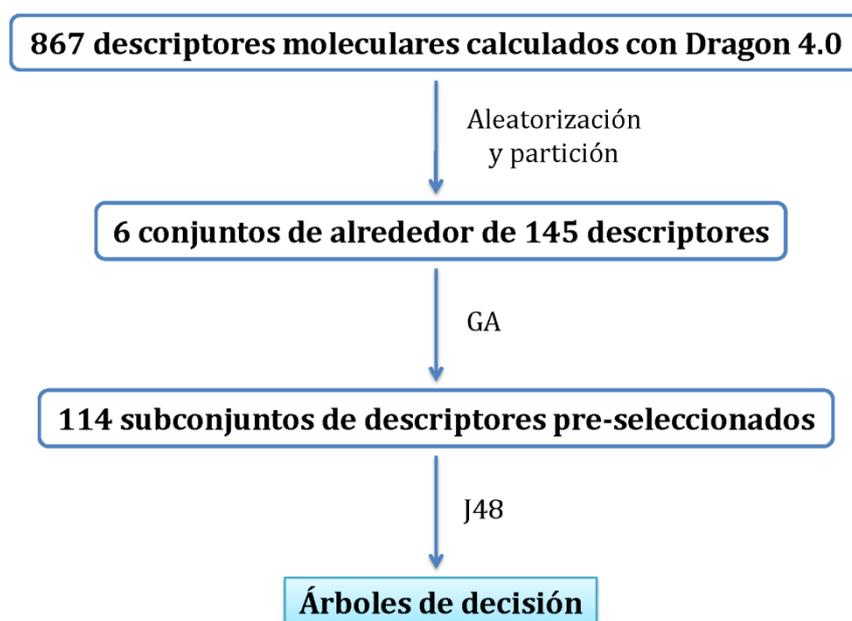
**Figura 4.11.** Esquema donde se muestra el método de poda en línea considerando un número mínimo de compuestos permitido por hoja de 10. Se puede observar que al reemplazarse el nodo por la hoja terminal quedan 5 compuestos mal clasificados. "x", "y", "z" y "j" representan el valor de corte del descriptor de cada nodo.

Para el evaluador de la capacidad clasificadora de un subconjunto de los GA, sólo se aplicó poda en línea variando sistemáticamente el parámetro M de 2 a 20. Se realizaron 4 corridas consecutivas donde el resultado de una corrida se utilizó como conjunto de descriptores de entrada para la siguiente, hasta alcanzar la convergencia, donde no se observó una mejora en la respuesta. De esta manera para cada uno de los 6 conjuntos originales de descriptores aleatorizados se obtuvieron 19 soluciones diferentes (subconjuntos de descriptores pre-seleccionados por GA), es decir 114 subconjuntos de descriptores en total.

Los modelos finales se obtuvieron aplicando el algoritmo J48 sobre cada uno de los 114 subconjuntos de descriptores obtenidos por GA. Para ello se aplicó la siguiente metodología:

- 1) el parámetro M se varió sistemáticamente de 2 a 20. Para cada valor de M, el parámetro C (factor de confianza) se varió sistemáticamente entre 0.001 y 0.5 cada 0.01 y para cada árbol final así obtenido se aplicó la elevación de subárboles.
- 2) Por otro lado sobre cada uno de los 114 subconjuntos se aplicó el método de poda de error reducido donde se varió el parámetro N (número de particiones del conjunto de entrenamiento) sistemáticamente entre 5 y 15.

En la Figura 4.12 se presenta un esquema del procedimiento de modelado seguido en esta etapa.



**Figura 4.12.** Representación esquemática simplificada de los pasos seguidos para la obtención de los modelos no lineales.

### 4.5 Evaluación del desempeño de los modelos

Para la evaluación y comparación del desempeño de todos los modelos generados se utilizaron las curvas ROC (por el inglés, *Receiver Operating Characteristic*) [77] que son representaciones gráficas de la sensibilidad del modelo ( $Se$ , la tasa de verdaderos positivos) frente a 1 menos la especificidad ( $Sp$ , la tasa de verdaderos negativos), que se definen por las siguientes expresiones:

$$Se = \frac{VP}{VP + FN}$$

$$Sp = \frac{VN}{VN + FP}$$

donde  $FN$  denota los falsos negativos y  $FP$  representa los falsos positivos. En este caso particular se buscan compuestos que no sean reconocidos por la BCRP (no sustratos) por lo que resultan ser las instancias positivas, mientras que se pretende descartar los compuestos sustratos de la BCRP, que por lo tanto serán considerados como las instancias negativas. De este modo, las expresiones anteriores pueden reescribirse de la siguiente manera:

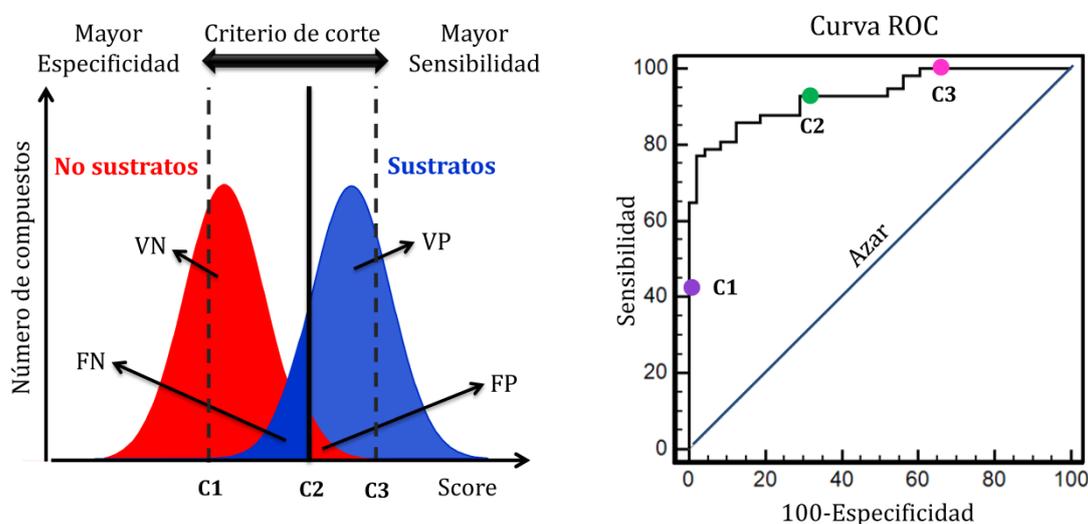
$$Se = \frac{\text{Verdaderos No sustratos}}{\text{Verdaderos No sustratos} + \text{Falsos Sustratos}}$$

$$Sp = \frac{\text{Verdaderos Sustratos}}{\text{Verdaderos Sustratos} + \text{Falsos No sustratos}}$$

donde un  $FN$  es un no sustrato predicho como sustrato por el modelo y un  $FP$  es un sustrato predicho como no sustrato.

El área bajo la curva ROC (AUC ROC) constituye una medida valiosa para evaluar si el modelo se comporta significativamente mejor que una clasificación al azar o que otro modelo [77]. Un modelo ideal presentará un AUC ROC de 1 (equivalente a una clasificación perfecta), mientras que la clasificación al azar se representa por una línea de pendiente 1 y corresponde a un AUC de 0.5. Al

modificar la selección del valor de corte entre ambas categorías, la  $Se$  y la  $Sp$  evolucionarán de manera opuesta. En consecuencia, no es posible optimizar ambos parámetros simultáneamente, y se debe recurrir a una solución intermedia [77]. Equilibrar la tasa de  $FP$  y de  $FN$  depende de consideraciones pragmáticas que deben ser juzgadas por el investigador [78]. En este caso se adoptó una actitud conservadora y se priorizó el desarrollo de modelos altamente específicos (con alta  $Sp$ ), es decir, modelos capaces de descartar prácticamente la totalidad de los sustratos de la BCRP. Por lo tanto, se le dio privilegio a la  $Sp$  sobre la  $Se$ : a riesgo de perder compuestos valiosos al aplicar los modelos generados en futuras campañas de cribado virtual, se eligió evitar la adquisición o síntesis de un fármaco candidato que, una vez evaluado en los ensayos farmacológicos, demostrará ser un  $FP$  (un compuesto que fue predicho como no sustrato pero que en realidad es transportado por la BCRP). En la Figura 4.13 se muestra un esquema de la construcción de una curva ROC.



**Figura 4.13.** Esquema de la construcción de una curva ROC.

Las curvas ROC se construyeron utilizando el programa MedCalc (MedCalc software, 2011) y para la comparación estadística de dos AUC ROC se utilizó el

método no paramétrico desarrollado por DeLong y col. [79] para el cálculo del error estándar de cada AUC y luego se calculó el z-estadístico con el fin de obtener el correspondiente valor de p [80].

## 4.6 Validación *in silico* de los modelos

Para la validación *in silico* de todos los modelos obtenidos se utilizaron técnicas de validación estándar de modo de evaluar la robustez y la capacidad predictiva.

### 4.6.1 Validación interna

La validación interna se realizó mediante las técnicas de aleatorización de Fisher, validación cruzada estratificada LGO (del inglés, *stratified leave-group-out-LGO- cross-validation*) y LOO (del inglés, *leave-one-out -LOO- cross-validation*). La validación cruzada LGO implica remover grupos de compuestos del conjunto de entrenamiento, regenerar el modelo con los compuestos remanentes y utilizar este nuevo modelo para predecir la clase de los compuestos removidos. El término estratificada alude a que se eligen las particiones de compuestos removidos de manera tal que la media de la variable dependiente es aproximadamente la misma en todas las particiones; en este caso esto se traduce en que cada partición removida tiene aproximadamente la misma cantidad de sustratos y no sustratos. Para los primeros modelos lineales se aplicó la técnica LGO donde cada ronda se llevó a cabo extrayendo al azar 10 compuestos del conjunto de entrenamiento, proceso que fue repetido 50 veces, comprobando que todos los compuestos del conjunto de entrenamiento hayan sido retirados en al menos una ronda de LGO. Para el caso de los modelos no lineales (árboles de decisión) el proceso fue repetido 150 veces. Para la segunda serie de modelos lineales se aplicó la técnica LOO repitiendo el cálculo 164 veces de modo de remover todos los compuestos del conjunto de entrenamiento una vez.

Por otro lado, la aleatorización implica precisamente aleatorizar los valores de la variable dependiente en el conjunto de entrenamiento, aboliendo por lo tanto cualquier relación que pudiera haber entre estructura molecular de los compuestos y la variable dependiente modelada, y generar nuevos modelos para evaluar si existe probabilidad de correlación azarosa entre las variables independientes y la variable dependiente. Se espera que los modelos obtenidos por aleatorización muestren un desempeño marcadamente inferior al del modelo auténtico. En este caso se construyeron 50 modelos aleatorios para cada caso.

#### 4.6.2 Validación externa

La validación externa se realizó primeramente sobre el conjunto de prueba de 98 compuestos generado especialmente a tal fin durante la etapa de partición del conjunto de datos. La clase de los compuestos del conjunto de prueba ha sido experimentalmente determinada pero no han sido utilizados durante el proceso de entrenamiento del modelo, por lo que se consideran independientes o externos al mismo.

Sin embargo, los experimentos conducidos por Truchon y Bayly sugieren que, cuando se utilizan métricas de enriquecimiento como las que surgen de las curvas ROC para evaluar los modelos sobre un conjunto de compuestos demasiado pequeño los valores calculados presentan un mayor error; el mismo se reduce y converge a un valor constante para bases de datos de mayor tamaño [81]. El otro problema que aparece cuando se trabaja con una elevada proporción de instancias (compuestos) positivas es lo que se conoce como "efecto de saturación": una vez que éstas instancias saturan la primera parte de la lista ordenada según el método de rankeado propuesto, las métricas de enriquecimiento no pueden mejorar; este efecto se atenúa cuando la relación instancias positivas/instancias negativas del conjunto de datos sometido a cribado es mucho menor a 1. Por este motivo, si un modelo *in silico* será aplicado para analizar grandes bibliotecas de compuestos químicos donde muy pocas instancias positivas están dispersas entre una gran cantidad de instancias negativas, el conjunto de prueba original de 98 compuestos

resulta insuficiente para evaluar la utilidad de los modelos en una aplicación de cribado real y es conveniente estimar ésta de manera más realista, examinando el comportamiento de los modelos al ser aplicados a una base de datos en la que un número pequeño de instancias positivas (no sustratos) se encuentre inmerso en un alto número de instancias negativas (sustratos o presuntos sustratos). Por lo tanto, para realizar una evaluación más rigurosa de la capacidad de enriquecimiento de los modelos, se construyeron dos bibliotecas químicas más grandes que el conjunto de prueba original de 98 compuestos, en las cuales la relación entre no sustratos/sustratos fuera menor a 0.05, a fin de realizar con ellas campañas piloto de cribado virtual.

#### 4.6.2.1 Biblioteca simulada

En primer lugar se construyó una biblioteca simulada de 577 compuestos donde el conjunto de prueba original de 98 compuestos (27 no sustratos y 71 sustratos) se dispersó entre 479 sustratos putativos que actúan como señuelos. Dado que el número de sustratos conocidos de la BCRP es limitado, tales sustratos putativos son los sustratos de la BCRP de otras especies que fueron apartados durante la construcción del conjunto de datos final, los sustratos que fueron evaluados en BCRP humana pero expresada en células no humanas que también fueron apartados, y compuestos obtenidos de las bases de datos ZINC [82] y PubChem [83] con un valor de similitud mayor a 0.75 con respecto a los sustratos del conjunto de prueba. Por lo tanto, esta biblioteca simulada contiene 27 no sustratos conocidos entre 550 sustratos conocidos o putativos. Se debe tener en cuenta que algunos de estos sustratos putativos podrían ser en realidad no sustratos; por lo tanto, el verdadero rendimiento de los modelos puede ser incluso mejor al obtenido a través de este experimento simulado.

#### 4.6.2.2 Biblioteca de señuelos DUD-E

Por otro lado y como un desafío final para los modelos, se construyó una biblioteca de mayor tamaño y diversidad estructural que la biblioteca simulada del

punto anterior, denominada biblioteca de señuelos DUD-E (por el inglés, *Enhanced Directory of Useful Decoys* [84-85]). La misma contiene 1346 compuestos (1248 señuelos sumados al conjunto de prueba original) donde cada señuelo es fisicoquímicamente similar pero topológicamente diferente a su correspondiente no sustrato. Para ello se utilizó la herramienta de generación automática de señuelos disponible en la web [86]. Brevemente, los señuelos son emparejados apropiadamente a los no sustratos utilizando como descriptores el peso molecular, la transformación logarítmica estimada del coeficiente de reparto octanol-agua (miLogP), número de enlaces rotables, recuento de enlaces aceptores de hidrógeno, recuento de enlaces dadores de hidrógeno y la carga molecular neta. Se generaron todos los estados de protonación de los no sustratos en el rango de pH 6-8. Por sobre todas las formas protonadas de un no sustrato dado, se conservaron sólo aquellas con un conjunto único de las seis propiedades fisicoquímicas consideradas. Para cada uno de estos seis conjuntos de propiedades únicas, se generaron cerca de 50 señuelos. Para lograr esto, se seleccionaron de ZINC [82] subconjuntos de señuelos utilizando un protocolo dinámico que estrecha o ensancha la ventanas de búsqueda asociada a cada una de las 6 propiedades fisicoquímicas utilizadas con el objetivo de lograr entre 3000-9000 potenciales señuelos para cada no sustrato empleado como *query*. Finalmente se generaron las huellas dactilares o *fingerprints* ECFP4 para los no sustratos y sus potenciales señuelos. Los señuelos son ordenados según su máximo coeficiente de Tanimoto para cada no sustrato y se conserva el 25% más disímil.

#### **4.7 Combinación de modelos: fusión de datos**

Se recurrió a la fusión de datos o aprendizaje por ensamblado para lograr, mediante consenso, predicciones más robustas [87]. La suposición que hay detrás del uso de la fusión de datos tiene un fundamento estadístico según el cual cuantas más veces se recupera una molécula mediante métodos o modelos independientes, mayor es la probabilidad de que la misma reúna las características de interés. Se ha sugerido, por otro lado, que el aprendizaje por ensamblado podría eliminar la

necesidad de estimar el dominio de aplicación de los modelos utilizados en el VS, asegurando una mayor capacidad de cobertura del espacio químico [88].

Teniendo en cuenta que la BCRP se caracteriza por una amplia especificidad de sustrato lo cual se debe muy probablemente a la existencia de múltiples sitios de unión en la proteína [89-94], se optó por la búsqueda de combinaciones de modelos que proporcionen la menor tasa de *FP* (falsos no sustratos). La poliespecificidad de sustrato y por ende su amplia diversidad estructural indica que podría ser difícil obtener un modelo único que sea capaz de identificar el conjunto completo de sustratos de la BCRP, como sugieren otros estudios de modelado enfocados en la predicción de otros transportadores MDR-ABC con características similares [95-100] y como podrá apreciarse luego en nuestros resultados.

#### *4.7.1 Combinación de los modelos lineales*

La primera serie de modelos lineales se combinaron por la sencilla estrategia de buscar exhaustivamente todas las posibles combinaciones de dos modelos. Se utilizaron como esquemas de fusión de datos el valor máximo (operador MAX) y la media (promedio) entre los dos valores (scores) proporcionados para cada compuesto por los dos clasificadores independientes que componen el conjunto.

Teniendo en cuenta que la combinación de pocos pero buenos modelos (combinación selectiva) puede proporcionar una mejor precisión y generalización que la combinación de todos los modelos disponibles [87, 101-102], la segunda serie de modelos lineales se combinaron de la siguiente manera: a) combinación de los scores de los 10 modelos individuales que presentaron la mejor AUC ROC para el conjunto de prueba y, b) combinación de los scores de todos los modelos individuales que mostraron un porcentaje de buenas clasificaciones global sobre el conjunto de prueba por encima del 70%. Se utilizaron cinco esquemas de fusión de datos: el operador MAX, el valor mínimo (operador MIN), el promedio, el ranking promedio y el voto promedio.

El ranking promedio consiste en generar un ranking de los compuestos en función del score que le asigna cada modelo y luego promediárselo; para esto se ordena el score de mayor a menor, se le asigna un número al ordenamiento y luego se calcula el promedio del ordenamiento asignado por cada uno de los modelos combinados.

El voto promedio se calcula de la siguiente manera: el voto obtenido por el compuesto  $j$ -ésimo en el modelo  $i$ -ésimo es igual a  $\max(0, \text{int}(11 - \text{rank}_{ij} / 0.02NDB))$ , donde  $\text{rank}_{ij}$  es el ranking del compuesto  $j$  según el modelo  $i$ , y  $NDB$  es el número de compuestos total de la base de datos evaluada. Este procedimiento otorga 10 votos a favor del primer 2% de los compuestos rankeados, 9 votos para el próximo 2%, y así sucesivamente. Para los compuestos rankeados entre el rango de 18-20%, se da 1 voto. Los compuestos del 80% inferior de la lista de clasificación no reciben ningún voto [53].

#### *4.7.2 Combinación de los modelos no lineales*

Para el caso de los árboles de decisión se realizó una combinación selectiva de los 12 mejores modelos individuales que presentaron la mejor AUC ROC tanto para el conjunto de prueba como para la biblioteca simulada. Se utilizaron los mismos cinco esquemas de fusión de datos que en el caso de la segunda serie de modelos lineales.

### **4.8 Selección de los mejores modelos y/o combinaciones**

Para elegir los mejores modelos individuales y las mejores combinaciones de modelos generados se utilizaron los siguientes criterios:

- 1) observación del % de buenas clasificaciones en el conjunto de entrenamiento, el conjunto de prueba, la biblioteca simulada y la biblioteca de señuelos DUD-E, con especial énfasis en los tres últimos.

- 2) observación del AUC ROC calculada tanto para el conjunto de entrenamiento, el conjunto de prueba, la biblioteca simulada y la biblioteca de señuelos DUD-E, con especial énfasis en los tres últimos.
- 3) observación del principio de parsimonia: entre dos modelos (y/o combinaciones) con similar comportamiento se eligió el más sencillo, es decir, el que incluyera menor número de descriptores (o de modelos).

#### ***4.9 Validación experimental de la mejor combinación de modelos obtenida***

Además de la ya descrita validación *in silico* de los modelos y combinaciones de los mismos, se realizó la validación experimental de la mejor combinación de modelos obtenida de modo de aportar mayor evidencia sobre la confiabilidad de las predicciones efectuadas.

##### *4.9.1 Selección de los candidatos a evaluar experimentalmente*

Se aplicó la mejor combinación de modelos obtenida sobre una pequeña base de datos conformada por 21 compuestos sintetizados o adquiridos por el Laboratorio de Investigación y Desarrollo de Bioactivos (LIDeB) de la Facultad de Ciencias Exactas, UNLP que presentaron las siguientes características:

- 1) Actividad anticonvulsiva comprobada
- 2) Absorción al UV, para su posterior determinación analítica por HPLC-UV.

Las estructuras químicas de los compuestos que conforman esta pequeña base de datos se muestran en el Anexo I.

Los compuestos a evaluar se seleccionaron dentro de aquellos clasificados como no sustrato de la BCRP por la combinación de modelos y que cayeron dentro del dominio de aplicación de los modelos que constituyen la combinación utilizada. Las predicciones (clasificaciones, en nuestro caso) de los modelos QSAR tendrán validez en el espacio químico definido por el conjunto de compuestos utilizados

para generar los modelos (conjunto de entrenamiento), mientras que todas aquellas predicciones de compuestos que caigan fuera de este espacio se considerarán extrapolaciones del modelo [103]. Ese espacio, dentro del cual las predicciones de un modelo QSAR son confiables, se denomina dominio de aplicación. Existen varios enfoques para calcular el dominio de aplicación de los modelos QSAR [103-104]; una de las formas más sencillas es empleando la Regla de la Palanca (en inglés, *leverage*): se determinan los valores de palanca ( $h_i$ ) de cada compuesto de la base de datos a evaluar, siendo  $x_i$  el vector de los descriptores del compuesto  $i$  y  $X$  la matriz del modelo obtenida a partir de los valores de los descriptores del conjunto de entrenamiento [105].

$$h_i = x_i^T (X^T X)^{-1} x_i (i = 1, \dots, n)$$

Un compuesto estará dentro del dominio de aplicabilidad del modelo cuando presente un valor de  $h$  inferior al umbral ( $h^*$ ), definido generalmente como  $h^* = 3f/n$ , donde  $f$  es el número de descriptores incluidos en el modelo y  $n$  el número de compuestos que conforman el conjunto de entrenamiento. Como ya se mencionó en la sección 4.7, la propia combinación de modelos provee cierta independencia de la determinación del dominio de aplicación de cada modelo individual; aún así, hemos adoptado un enfoque conservador calculando de todos modos el dominio de aplicación para cada predicción.

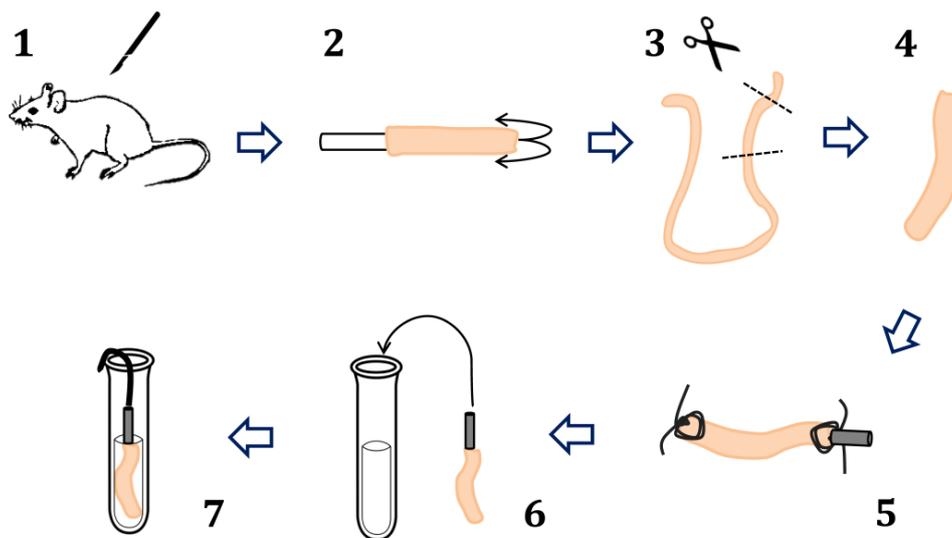
#### 4.9.2 Evaluación experimental de los compuestos seleccionados

Se utilizó el modelo *ex vivo* de saco intestinal evertido de rata para evaluar el transporte mediado por la BCRP [106-107]. La principal ventaja de esta técnica con respecto a los ensayos *in vitro* es que con frecuencia los resultados obtenidos mediante esta técnica *ex vivo* han estado de concordancia con los estudios *in vivo* [108-112]. Otras ventajas de este modelo *ex vivo* para estudiar los mecanismos y la cinética de transporte de un fármaco son el área de superficie relativamente

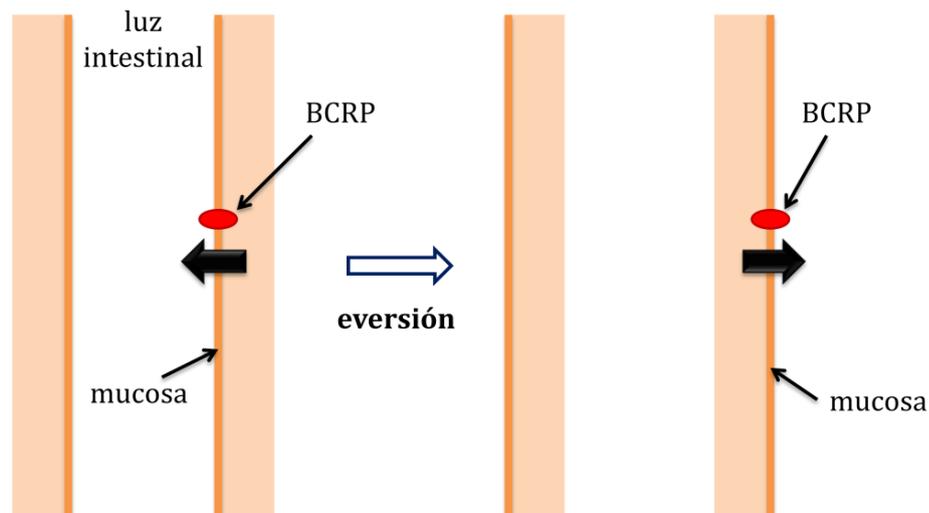
grande disponible para la absorción y la presencia de mucus [113], dos características importantes cuando se desea predecir la biodisponibilidad oral de un fármaco. Si bien el principal interés, en este caso, es descartar la afinidad de los candidatos anticonvulsivos por la BCRP dada la regulación hacia arriba de la misma a nivel del foco epiléptico, no está demás reunir información sobre la posible influencia de este transportador sobre la biodisponibilidad oral de los fármacos estudiados. Por otra parte, la viabilidad del tejido (un máximo de 2 hs.) es una de las principales limitaciones [114]. Se han empleado diferentes animales para los experimentos de saco intestinal evertido siendo la rata el animal más utilizado [107, 114-116]. La BCRP está presente en todos los segmentos del intestino delgado de la rata, desde el duodeno y la parte proximal del yeyuno hasta el yeyuno distal y la válvula ileocecal, incluyendo el íleon. Los mayores niveles de expresión de la BCRP se dan en el segmento que incluye el yeyuno distal y el íleon, y por lo tanto se utilizó esta porción del intestino para estudiar la posible interacción de los fármacos anticonvulsivos por la BCRP [107, 117].

Brevemente, para los ensayos se utilizaron ratas macho Wistar (280-320 gr. de peso corporal) que fueron mantenidas bajo un ciclo de luz/oscuridad de 12/12 hs., a temperatura ambiente controlada, con comida y agua *ad libitum*. Los experimentos se llevaron a cabo de conformidad con la Guía para el Cuidado y Uso de Animales de Laboratorio del Consejo Superior de Investigaciones Científicas (EE.UU., 1996) y también de acuerdo con la disposición N° 6344/96 de la Administración Nacional de Medicamentos, Alimentos y Tecnología Médica (ANMAT). En cada ensayo las ratas fueron anestesiadas con uretano (1,2 gr/kg, inyección ip), se efectuó la apertura del abdomen y posterior retiro del extremo distal del intestino delgado, el cual fue lavado y perfundido con solución de Krebs fría burbujeada con 95% O<sub>2</sub>/5% de CO<sub>2</sub> (NaCl 118 mM; KCl 4.7 mM; MgCl<sub>2</sub> 1.2 mM; NaH<sub>2</sub>PO<sub>4</sub> 1.0 mM; CaCl<sub>2</sub> 2.6 mM; NaHCO<sub>3</sub> 25.0 mM; glucosa 11.1 mM; etilendiaminotetraacetato disódico (EDTA Na<sub>2</sub>) 0.004 mM; ácido ascórbico 0.11 mM; pH final 7,4). El intestino fue evertido suavemente sobre una varilla de vidrio, se dividió en cuatro segmentos de 5-7 cm cada uno. Para generar los sacos, uno de los extremos de cada segmento de intestino se ligó firmemente atándolo con hilo, mientras que el otro extremo se ató también firmemente con hilo a una cánula intravenosa corta por donde se administraron las soluciones. Un alambre delgado

de acero introducido en el casquillo de la cánula actuó como un gancho que permitió colgar el saco intestinal evertido en el recipiente donde se realizó el ensayo. Los sacos se sumergieron en 5 ml de buffer Krebs precalentado a 37 °C bajo burbujeo con 95% O<sub>2</sub>/5% de CO<sub>2</sub> y se los dejó estabilizar 15 minutos luego de los cuales se inició el ensayo introduciendo la solución de la droga a evaluar dentro de cada saco (lado serosal) con la ayuda de una jeringa. Este punto de tiempo se consideró el tiempo 0. En el caso de evaluar inhibición del transportador, el inhibidor se añadió al buffer 30 minutos antes de la adición de la solución de la droga en el interior de los sacos de intestino para permitir que el inhibidor entre en contacto con el transportador. El transporte de la droga a través del intestino desde la serosa a la superficie mucosa se evaluó mediante la toma de muestras del medio (150 µl con reposición) cada 5 minutos durante 30 minutos, tiempo en que se dio por finalizado el ensayo. En la Figura 4.14 se muestra un esquema del procedimiento llevado a cabo durante el experimento de saco intestinal evertido de rata, mientras que en la Figura 4.15 se muestra la ubicación de la BCRP antes y después de la eversión del tejido intestinal.



**Figura 4.14.** Se muestra un esquema del procedimiento llevado a cabo durante el ensayo de saco intestinal evertido de rata. Paso 1: luego de anestesiarse la rata se efectúa la apertura del abdomen y posterior retiro del extremo distal del intestino delgado. Pasos 2, 3 y 4: el intestino es evertido suavemente sobre una varilla de vidrio y se divide en segmentos. Paso 5: para generar los sacos, uno de los extremos de cada segmento de intestino se liga firmemente atándolo con hilo, mientras que el otro extremo se ata también firmemente con hilo a una cánula intravenosa corta por donde se administran las soluciones. Paso 6 y 7: un alambre delgado de acero introducido en el casquillo de la cánula actúa como un gancho que permite colgar el saco intestinal evertido en el recipiente donde se realiza el ensayo.



**Figura 4.15.** La BCRP se encuentra localizada en la luz del intestino (es decir, en la mucosa), por lo que al evertir el saco intestinal la mucosa queda expuesta hacia el exterior, y por este motivo se espera que los sustratos de la BCRP introducidos en el interior del saco (el lado serosal), sean transportados hacia el exterior, es decir desde la serosa a la superficie mucosa.

Se utilizó nitrofurantoína (NF) como sustrato conocido [118] y Ko143 como inhibidor específico de la BCRP [119] para comprobar el transporte activo mediado por el transportador en el intestino de la rata. La NF fue cordialmente donada por Laboratorios Bagó SA y el Ko143 fue adquirido de Sigma-Aldrich Argentina. Se evaluaron tres concentraciones de NF (1, 10 y 100  $\mu\text{M}$ ) de modo de establecer la concentración óptima de trabajo a utilizarse como control positivo a la cual se observe analíticamente el transporte. Una vez establecida la concentración de NF, se evaluaron dos concentraciones del inhibidor (10  $\mu\text{M}$  y 50  $\mu\text{M}$ ) de modo de establecer la concentración óptima a la cual se observe la inhibición del transporte esperada.

La evaluación de las drogas seleccionadas se realizó mediante dos protocolos:

**Protocolo 1:** evaluación de la potencial interacción con el transportador midiendo efectos de cada droga en el transporte de NF. Para ello se utilizó el protocolo establecido para los inhibidores; es decir, se añadió cada droga al buffer 15 minutos antes de comenzar el ensayo. Se utilizó la concentración de NF ya

establecida y se evaluaron dos concentraciones de cada droga correspondientes a la concentración óptima del inhibidor y de NF respectivamente.

**Protocolo 2:** evaluación del transporte de aquellas drogas que no evidenciaron interacción estadísticamente significativa con la BCRP en el protocolo 1 (es decir, aquellas drogas que no presentaron efecto sobre el transporte de NF y por lo tanto tengan altas probabilidades de ser no sustrato de BCRP) en presencia y ausencia de inhibidor. Aquí debe aclararse que sólo fue posible la evaluación de aquellas drogas que estuvieron dentro del límite de detección del método analítico de cuantificación disponible (HPLC-UV). Cada droga se evaluó individualmente siguiendo el protocolo establecido para NF; es decir, se inyectó la solución de la droga dentro del saco y se siguió el mismo esquema de muestreo. La concentración evaluada de cada droga y de inhibidor corresponden a la concentración óptima establecida para NF y el inhibidor respectivamente. También se realizó un control a 4°C para descartar o evidenciar un posible transporte activo de las drogas independiente de la BCRP.

### *4.9.3 Análisis cuantitativo de las muestras por cromatografía líquida de alta resolución (HPLC)*

Se desarrolló y validó un método analítico por HPLC-UV para la cuantificación de las todas las drogas evaluadas en las matrices provenientes de los estudios *ex vivo*. Sobre 150 µl de muestra se realizó la precipitación de las proteínas mediante el agregado de 150 µl de acetonitrilo y posterior centrifugación. Se utilizó un estándar interno adecuadamente seleccionado en cada caso. Todas las muestras fueron cuantificadas utilizando un equipo de UHPLC Dionex Ultimate 3000 (Thermo Scientific, Dionex, Sunnyvale, California, USA) equipado con un detector con arreglo de diodos. La fase estacionaria consistió en una columna Luna RP18 (5 µm, 150 × 4,6 mm) (Phenomenex, Torrance, CA, USA), operada a temperatura ambiente. En todos los casos la fase móvil estuvo compuesta por:

A. Buffer  $\text{KH}_2\text{PO}_4$  20 mM, ajustado a pH 2,5 con  $\text{H}_3\text{PO}_4$  y

#### B. Metanol

Las condiciones particulares de la fase móvil para cada droga serán detalladas en el siguiente capítulo. Para la cuantificación de cada droga se verificó la linealidad, especificidad y precisión en el intervalo concentraciones de trabajo.

#### *4.9.4 Análisis estadístico de los resultados experimentales*

Para el análisis estadístico de los resultados se verificaron en primer lugar los supuestos necesarios para realizar un ANOVA: aleatoriedad de los residuos, normalidad de los datos y homogeneidad de varianzas. En los casos en que fue necesario se aplicó la transformación logarítmica de los datos. Finalmente se aplicó, según el caso, el test t de Student o el test factorial de análisis de varianza (ANOVA monofactorial).

---

**Referencias Capítulo IV**

- [1] Zhang L, Balimane PV, Johnson SR, Chong S. Development of an in silico model for predicting efflux substrates in Caco-2 cells. *Int J Pharm.* 2007;343(1-2):98-105.
- [2] Ultra-Noguera D, Mangas-Sanjuan V, Centelles-Sanguesa A, Gonzalez-Garcia I, Sanchez-Castano G, Gonzalez-Alvarez M, et al. Variability of permeability estimation from different protocols of subculture and transport experiments in cell monolayers. *J Pharmacol Toxicol Methods.* 2015;71:21-32.
- [3] Allen JD, Jackson SC, Schinkel AH. A mutation hot spot in the Bcrp1 (Abcg2) multidrug transporter in mouse cell lines selected for Doxorubicin resistance. *Cancer Res.* 2002;62(8):2294-9.
- [4] Robey RW, Honjo Y, Morisaki K, Nadjem TA, Runge S, Risbood M, et al. Mutations at amino-acid 482 in the ABCG2 gene affect substrate and antagonist specificity. *Br J Cancer.* 2003;89(10):1971-8.
- [5] Polgar O, Robey RW, Bates SE. ABCG2: structure, function and role in drug response. *Expert Opin Drug Metab Toxicol.* 2008;4(1):1-15.
- [6] Pozza A, Perez-Victoria JM, Sardo A, Ahmed-Belkacem A, Di Pietro A. Purification of breast cancer resistance protein ABCG2 and role of arginine-482. *Cell Mol Life Sci.* 2006;63(16):1912-22.
- [7] Janvilisri T, Shahi S, Venter H, Balakrishnan L, van Veen HW. Arginine-482 is not essential for transport of antibiotics, primary bile acids and unconjugated sterols by the human breast cancer resistance protein (ABCG2). *Biochem J.* 2005;385(Pt 2):419-26.
- [8] Ejendal KF, Diop NK, Schweiger LC, Hrycyna CA. The nature of amino acid 482 of human ABCG2 affects substrate transport and ATP hydrolysis but not substrate binding. *Protein Sci.* 2006;15(7):1597-607.
- [9] Eddabra L, Wenner T, El Btaouri H, Baranek T, Madoulet C, Cornillet-Lefebvre P, et al. Arginine 482 to glycine mutation in ABCG2/BCRP increases etoposide transport and resistance to the drug in HEK-293 cells. *Oncol Rep.* 2012;27(1):232-7.
- [10] Kathawala RJ, Gupta P, Ashby CR, Jr., Chen ZS. The modulation of ABC transporter-mediated multidrug resistance in cancer: a review of the past decade. *Drug Resist Updat.* 2015;18:1-17.

- [11] Ni Z, Bikadi Z, Rosenberg MF, Mao Q. Structure and function of the human breast cancer resistance protein (BCRP/ABCG2). *Curr Drug Metab.* 2010;11(7):603-17.
- [12] Jani M, Ambrus C, Magnan R, Jakab KT, Beery E, Zolnerciks JK, et al. Structure and function of BCRP, a broad specificity transporter of xenobiotics and endobiotics. *Arch Toxicol.* 2014;88(6):1205-48.
- [13] Shafran A, Ifergan I, Bram E, Jansen G, Kathmann I, Peters GJ, et al. ABCG2 harboring the Gly482 mutation confers high-level resistance to various hydrophilic antifolates. *Cancer Res.* 2005;65(18):8414-22.
- [14] Özvegy-Laczka C, Köblös G, Sarkadi B, Váradi A. Single amino acid (482) variants of the ABCG2 multidrug transporter: major differences in transport capacity and substrate recognition. *Biochimica et Biophysica Acta.* 2005;1668(1):53-63.
- [15] Honjo Y, Hrycyna CA, Yan QW, Medina-Perez WY, Robey RW, van de Laar A, et al. Acquired mutations in the MXR/BCRP/ABCP gene alter substrate specificity in MXR/BCRP/ABCP-overexpressing cells. *Cancer Res.* 2001;61(18):6635-9.
- [16] Gal Z, Hegedus C, Szakacs G, Varadi A, Sarkadi B, Ozvegy-Laczka C. Mutations of the central tyrosines of putative cholesterol recognition amino acid consensus (CRAC) sequences modify folding, activity, and sterol-sensing of the human ABCG2 multidrug transporter. *Biochim Biophys Acta.* 2015;1848(2):477-87.
- [17] Ozvegy-Laczka C, Koblos G, Sarkadi B, Varadi A. Single amino acid (482) variants of the ABCG2 multidrug transporter: major differences in transport capacity and substrate recognition. *Biochim Biophys Acta.* 2005;1668(1):53-63.
- [18] Cervenak J, Andrikovics H, Ozvegy-Laczka C, Tordai A, Nemet K, Varadi A, et al. The role of the human ABCG2 multidrug transporter and its variants in cancer therapy and toxicology. *Cancer Lett.* 2006;234(1):62-72.
- [19] Suvannasankha A, Minderman H, O'Loughlin KL, Nakanishi T, Ford LA, Greco WR, et al. Breast cancer resistance protein (BCRP/MXR/ABCG2) in adult acute lymphoblastic leukaemia: frequent expression and possible correlation with shorter disease-free survival. *Br J Haematol.* 2004;127(4):392-8.
- [20] Backstrom G, Taipalensuu J, Melhus H, Brandstrom H, Svensson AC, Artursson P, et al. Genetic variation in the ATP-binding cassette transporter gene ABCG2 (BCRP) in a Swedish population. *Eur J Pharm Sci.* 2003;18(5):359-64.

- [21] Honjo Y, Morisaki K, Huff LM, Robey RW, Hung J, Dean M, et al. Single-nucleotide polymorphism (SNP) analysis in the ABC half-transporter ABCG2 (MXR/BCRP/ABCP1). *Cancer Biol Ther.* 2002;1(6):696-702.
- [22] Itoda M, Saito Y, Shirao K, Minami H, Ohtsu A, Yoshida T, et al. Eight novel single nucleotide polymorphisms in ABCG2/BCRP in Japanese cancer patients administered irinotecan. *Drug Metab Pharmacokinet.* 2003;18(3):212-7.
- [23] Kobayashi D, Ieiri I, Hirota T, Takane H, Maegawa S, Kigawa J, et al. Functional assessment of ABCG2 (BCRP) gene polymorphisms to protein expression in human placenta. *Drug Metab Dispos.* 2005;33(1):94-101.
- [24] Mizuarai S, Aozasa N, Kotani H. Single nucleotide polymorphisms result in impaired membrane localization and reduced atpase activity in multidrug transporter ABCG2. *Int J Cancer.* 2004;109(2):238-46.
- [25] Suvannasankha A, Minderman H, O'Loughlin KL, Nakanishi T, Greco WR, Ross DD, et al. Breast cancer resistance protein (BCRP/MXR/ABCG2) in acute myeloid leukemia: discordance between expression and function. *Leukemia.* 2004;18(7):1252-7.
- [26] Zamber CP, Lamba JK, Yasuda K, Farnum J, Thummel K, Schuetz JD, et al. Natural allelic variants of breast cancer resistance protein (BCRP) and their relationship to BCRP expression in human intestine. *Pharmacogenetics.* 2003;13(1):19-28.
- [27] Robey RW, Steadman K, Polgar O, Bates SE. ABCG2-mediated transport of photosensitizers: potential impact on photodynamic therapy. *Cancer Biol Ther.* 2005;4(2):187-94.
- [28] Van Hattum AH, Schluper HM, Hausheer FH, Pinedo HM, Boven E. Novel camptothecin derivative BNP1350 in experimental human ovarian cancer: determination of efficacy and possible mechanisms of resistance. *Int J Cancer.* 2002;100(1):22-9.
- [29] Allen JD, Van Dort SC, Buitelaar M, van Tellingen O, Schinkel AH. Mouse breast cancer resistance protein (Bcrp1/Abcg2) mediates etoposide resistance and transport, but etoposide oral availability is limited primarily by P-glycoprotein. *Cancer Res.* 2003;63(6):1339-44.
- [30] de Wolf C, Jansen R, Yamaguchi H, de Haas M, van de Wetering K, Wijnholds J, et al. Contribution of the drug transporter ABCG2 (breast cancer resistance

protein) to resistance against anticancer nucleosides. *Mol Cancer Ther.* 2008;7(9):3092-102.

[31] Hirano M, Maeda K, Matsushima S, Nozaki Y, Kusuhara H, Sugiyama Y. Involvement of BCRP (ABCG2) in the biliary excretion of pitavastatin. *Mol Pharmacol.* 2005;68(3):800-7.

[32] Matsushima S, Maeda K, Kondo C, Hirano M, Sasaki M, Suzuki H, et al. Identification of the hepatic efflux transporters of organic anions using double-transfected Madin-Darby canine kidney II cells expressing human organic anion-transporting polypeptide 1B1 (OATP1B1)/multidrug resistance-associated protein 2, OATP1B1/multidrug resistance 1, and OATP1B1/breast cancer resistance protein. *J Pharmacol Exp Ther.* 2005;314(3):1059-67.

[33] Pavek P, Merino G, Wagenaar E, Bolscher E, Novotna M, Jonker JW, et al. Human breast cancer resistance protein: interactions with steroid drugs, hormones, the dietary carcinogen 2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine, and transport of cimetidine. *J Pharmacol Exp Ther.* 2005;312(1):144-52.

[34] Imai Y, Asada S, Tsukahara S, Ishikawa E, Tsuruo T, Sugimoto Y. Breast cancer resistance protein exports sulfated estrogens but not free estrogens. *Mol Pharmacol.* 2003;64(3):610-8.

[35] Mao Q, Unadkat JD. Role of the breast cancer resistance protein (ABCG2) in drug transport. *AAPS J.* 2005;7(1):E118-33.

[36] Zamek-Gliszczyński M, Hoffmaster KA, Humphreys JE, Tian X, Nezasa K, Brouwer KL. Differential involvement of Mrp2 (Abcc2) and Bcrp (Abcg2) in biliary excretion of 4-methylumbelliferyl glucuronide and sulfate in the rat. *J Pharmacol Exp Ther.* 2006;319(1):459-67.

[37] Mizuno N, Suzuki M, Kusuhara H, Suzuki H, Takeuchi K, Niwa T, et al. Impaired renal excretion of 6-hydroxy-5,7-dimethyl-2-methylamino-4-(3-pyridylmethyl) benzothiazole (E3040) sulfate in breast cancer resistance protein (BCRP1/ABCG2) knockout mice. *Drug Metab Dispos.* 2004;32(9):898-901.

[38] Zamek-Gliszczyński MJ, Nezasa K, Tian X, Kalvass JC, Patel NJ, Raub TJ, et al. The important role of Bcrp (Abcg2) in the biliary excretion of sulfate and glucuronide metabolites of acetaminophen, 4-methylumbelliferone, and harmol in mice. *Mol Pharmacol.* 2006;70(6):2127-33.

- [39] Suzuki M, Suzuki H, Sugimoto Y, Sugiyama Y. ABCG2 transports sulfated conjugates of steroids and xenobiotics. *J Biol Chem.* 2003;278(25):22644-9.
- [40] Chen Z-S, Robey RW, Belinsky MG, Shchaveleva I, Ren X-Q, Sugimoto Y, et al. Transport of Methotrexate, Methotrexate Polyglutamates, and 17 $\beta$ -Estradiol 17-( $\beta$ -D-glucuronide) by ABCG2: Effects of Acquired Mutations at R482 on Methotrexate Transport. *Cancer Res* 2003;63:4048–54.
- [41] Bakhsheshian J, Hall MD, Robey RW, Herrmann MA, Chen JQ, Bates SE, et al. Overlapping substrate and inhibitor specificity of human and murine ABCG2. *Drug Metab Dispos.* 2013;41(10):1805-12.
- [42] Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aided Mol Des.* 2003;17(2-4):241-53.
- [43] Golbraikh A, Tropsha A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Mol Divers.* 2002;5(4):231-43.
- [44] Kawabata T. Build-up algorithm for atomic correspondence between chemical structures. *J Chem Inf Model.* 2011;51(8):1775-87.
- [45] Talevi A, Castro EA, Bruno-Blanch LE. Recent studies on similarity measures and its applications to chemoinformatics and drug design. *Recent Trends on QSAR in the Pharmaceutical Perceptions: Bentham Science; 2012.*
- [46] Hariharan R, Janakiraman A, Nilakantan R, Singh B, Varghese S, Landrum G, et al. MultiMCS: a fast algorithm for the maximum common substructure problem on multiple molecules. *J Chem Inf Model.* 2011;51(4):788-806.
- [47] Hartigan JA. *Clustering Algorithms.* New York, NY, USA: John Wiley & Sons, Inc.; 1975.
- [48] Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. *Applied Statistics.* 1978;28(1):100-8.
- [49] Everitt BS, Landau S, Leese M, Stahl D. *Optimization Clustering Techniques.* Cluster Analysis: Wiley; 2011.
- [50] Mercader AG, Duchowicz PR, Fernandez FM, Castro EA. Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories. *J Chem Inf Model.* 2010;50(9):1542-8.

- [51] Caballero J, Fernández M. Linear and non-linear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural network. *J Mol Model*. 2006;12:168-81.
- [52] Talevi A, Bellera CL, Di Ianni M, Duchowicz PR, Bruno-Blanch LE, Castro EA. An integrated drug development approach applying topological descriptors. *Curr Comput Aided Drug Des*. 2012;8(3):172-81.
- [53] Zhang Q, Muegge I. Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring. *J Med Chem*. 2006;49(5):1536-48.
- [54] Bender A, Glen RC. A discussion of measures of enrichment in virtual screening: comparing the information content of descriptors with increasing levels of sophistication. *J Chem Inf Model*. 2005;45(5):1369-75.
- [55] Yousefinejad S, Hemmateenejad B. Chemometrics tools in QSAR/QSPR studies: A historical perspective. *Chemometrics and Intelligent Laboratory Systems*. 2015;149(Part B):177-204.
- [56] O'Brien SE, de Groot MJ. Greater than the sum of its parts: combining models for useful ADMET prediction. *J Med Chem*. 2005;48(4):1287-91.
- [57] Varnek A, Baskin I. Machine learning methods for property prediction in chemoinformatics: Quo Vadis? *J Chem Inf Model*. 2012;52(6):1413-37.
- [58] Talevi A, Castro EA, Bruno-Blanch LE. Towards accurate virtual screening technologies: sampling of chemical space and applicability domain estimation. *Chemical Information and Computational Challenges in the 21st Century*: Nova Publishers; 2012.
- [59] Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol*. 1995;48(12):1503-10.
- [60] Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49(12):1373-9.
- [61] Mercader AG, Duchowicz PR, Fernandez FM, Castro EA. Advances in the replacement and enhanced replacement method in QSAR and QSPR theories. *J Chem Inf Model*. 2011;51(7):1575-81.

- [62] Mercader AG, Duchowicz PR, Fernández FM, Castro EA. Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories. *Chemometrics and Intelligent Laboratory Systems*. 2008;92(2):138-44.
- [63] Duchowicz PR, Castro EA, Fernández FM. Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies. *MATCH Commun Math Comput Chem*. 2006;55(1):179-92.
- [64] Blower PE, Cross KP. Decision tree methods in pharmaceutical research. *Curr Top Med Chem*. 2006;6(1):31-9.
- [65] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. *the WEKA Data Mining Software: An Update*. *SIGKDD Explorations*. 2009;11(1):10-8.
- [66] Quinlan JR. *C4.5: Programs for Machine Learning*. San Francisco, USA: Morgan Kaufmann Publishers; 1993.
- [67] Holland J. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Oxford, England: U Michigan Press; 1975.
- [68] Devillers J. Genetic algorithms in computer-aided molecular design. In: Devillers J, editor. *Genetic Algorithms in Molecular Modeling*. New York: Academic Press; 1996. p. 1-34.
- [69] Leardi R. Genetic algorithms in chemometrics and chemistry: a review. *J Chemom* 2001;15(7):559-69.
- [70] Niazi A, Leardi R. Genetic algorithms in chemometrics. *J Chemometrics* 2012;26(6):345-51.
- [71] Lucasius CB, Kateman G. Understanding and using genetic algorithms. Part 1. Concepts, properties and context. *Chemom Intell Lab Syst*. 1993;19:1-33.
- [72] Shannon CE. A mathematical theory of communication. *Bell System Technical Journal*. 1948;27(3):379-423.
- [73] Quinlan JR. Discovering rules by induction from large collections of examples. In: Michie D, editor. *Expert Systems in the Micro Electronic Age*. Edinburgh: Edinburgh University Press; 1979. p. 168-201.
- [74] Li X, Claramunt C. A Spatial Entropy-Based Decision Tree for Classification of Geographical Information. *Transactions in GIS*. 2006;10(3):451-67.

- [75] Bhargava N, Sharma G, Bhargava R, Mathuria M. Decision Tree Analysis on J48 Algorithm for Data Mining. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2013;3(6):1114-9.
- [76] Drazin S, Montag M. Decision Tree analysis using Weka. *Machine Learning-Project II*, University of Miami. 2012:1-3.
- [77] Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem*. 2005;48(7):2534-47.
- [78] Hubbard R, Bayarri MJ. Confusion over measures of evidence (p's) versus errors ( $\alpha$ 's) in classical statistical testing. *American Statistician*. 2003;57(3):171-8.
- [79] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-45.
- [80] Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med*. 2013;4(2):627-35.
- [81] Truchon JF, Bayly CI. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J Chem Inf Model*. 2007;47(2):488-508.
- [82] Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model*. 2005;45(1):177-82.
- [83] [www.pubchem.ncbi.nlm.nih.gov](http://www.pubchem.ncbi.nlm.nih.gov).
- [84] Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. *J Med Chem*. 2006;49(23):6789-801.
- [85] Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem*. 2012;55(14):6582-94.
- [86] <http://dude.docking.org/generate>.
- [87] Li L, Hu Q, Wu X, Yu D. Exploration of classification confidence in ensemble learning. *Pattern Recogn*. 2014;47(9):3120-31.
- [88] Zhu H, Tropsha A, Fourches D, Varnek A, Papa E, Gramatica P, et al. Combinatorial QSAR modeling of chemical toxicants tested against *Tetrahymena pyriformis*. *J Chem Inf Model*. 2008;48(4):766-84.

- [89] Giri N, Agarwal S, Shaik N, Pan G, Chen Y, Elmquist WF. Substrate-dependent breast cancer resistance protein (Bcrp1/Abcg2)-mediated interactions: consideration of multiple binding sites in in vitro assay design. *Drug Metab Dispos.* 2009;37(3):560-70.
- [90] Nakagawa R, Hara Y, Arakawa H, Nishimura S, Komatani H. ABCG2 confers resistance to indolocarbazole compounds by ATP-dependent transport. *Biochem Biophys Res Commun.* 2002;299(4):669-75.
- [91] Takenaka K, Morgan JA, Scheffer GL, Adachi M, Stewart CF, Sun D, et al. Substrate overlap between Mrp4 and Abcg2/Bcrp affects purine analogue drug cytotoxicity and tissue distribution. *Cancer Res.* 2007;67(14):6965-72.
- [92] Mo W, Qi J, Zhang JT. Different roles of TM5, TM6, and ECL3 in the oligomerization and function of human ABCG2. *Biochemistry.* 2012;51(17):3634-41.
- [93] Hazai E, Bikadi Z. Homology modeling of breast cancer resistance protein (ABCG2). *J Struct Biol.* 2008;162(1):63-74.
- [94] Clark R, Kerr ID, Callaghan R. Multiple drugbinding sites on the R482G isoform of the ABCG2 transporter. *Br J Pharmacol.* 2006;149(5):506-15.
- [95] Penzotti JE, Lamb ML, Evensen E, Grootenhuis PD. A computational ensemble pharmacophore model for identifying substrates of P-glycoprotein. *J Med Chem.* 2002;45(9):1737-40.
- [96] Li WX, Li L, Eksterowicz J, Ling XB, Cardozo M. Significance analysis and multiple pharmacophore models for differentiating P-glycoprotein substrates. *J Chem Inf Model.* 2007;47(6):2429-38.
- [97] Svetnik V, Wang T, Tong C, Liaw A, Sheridan RP, Song Q. Boosting: an ensemble learning tool for compound classification and QSAR modeling. *J Chem Inf Model.* 2005;45(3):786-99.
- [98] Di Ianni M, Talevi A, Castro EA, Bruno-Blanch LE. Development of a highly specific ensemble of topological models for early identification of P-glycoprotein substrates. *Journal of Chemometrics.* 2011;25(6):313-22.
- [99] Leong MK, Chen HB, Shih YH. Prediction of promiscuous p-glycoprotein inhibition using a novel machine learning scheme. *PLoS One.* 2012;7(3):e33829.

- [100] Chen L, Li Y, Yu H, Zhang L, Hou T. Computational models for predicting substrates or inhibitors of P-glycoprotein. *Drug Discov Today*. 2012;17(7-8):343-51.
- [101] Zhouh Z, Wu J, Tang W. Ensembling neural networks: Many could be better than all. *Artif Intell*. 2002;137(1-2):239-63.
- [102] Sastry GM, Inakollu VS, Sherman W. Boosting virtual screening enrichments with data fusion: coalescing hits from two-dimensional fingerprints, shape, and docking. *J Chem Inf Model*. 2013;53(7):1531-42.
- [103] Gramatica P. Principles of QSAR models validation: internal and external. *QSAR Comb Sci*. 2007;26(5):694-701.
- [104] Dimitrov S, Dimitrova G, Pavlov T, Dimitrova N, Patlewicz G, Niemela J, et al. A stepwise approach for defining the applicability domain of SAR and QSAR models. *J Chem Inf Model*. 2005;45(4):839-49.
- [105] Tropsha A, Gramatica P, Gombar VK. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR & Combinatorial Science*. 2003;22(1):69-77.
- [106] Rong Z, Xu Y, Zhang C, Xiang D, Li X, Liu D. Evaluation of intestinal absorption of amtolmetin guacyl in rats: breast cancer resistant protein as a primary barrier of oral bioavailability. *Life Sci*. 2013;92(3):245-51.
- [107] Peroni RN, Di Gennaro SS, Hocht C, Chiappetta DA, Rubio MC, Sosnik A, et al. Efavirenz is a substrate and in turn modulates the expression of the efflux transporter ABCG2/BCRP in the gastrointestinal tract of the rat. *Biochem Pharmacol*. 2011;82(9):1227-33.
- [108] Yumoto R, Hamada S, Okada K, Kato Y, Ikehata M, Nagai J, et al. Effect of ursodeoxycholic acid treatment on the expression and function of multidrug resistance-associated protein 2 in rat intestine. *J Pharm Sci*. 2009;98(8):2822-31.
- [109] Patel JP, Korashy HM, El-Kadi AO, Brocks DR. Effect of bile and lipids on the stereoselective metabolism of halofantrine by rat everted-intestinal sacs. *Chirality*. 2010;22(2):275-83.
- [110] Ni L, Yu X, Yu Q, Chen X, Jia L. Effects of cyclosporine A and itraconazole on permeability, biliary excretion and pharmacokinetics of amlodipine. *Drug Metab Lett*. 2008;2(3):163-8.

- [111] Machavaram KK, Gundu J, Yamsani MR. Effect of ketoconazole and rifampicin on the pharmacokinetics of ranitidine in healthy human volunteers: a possible role of P-glycoprotein. *Drug Metabol Drug Interact.* 2006;22(1):47-65.
- [112] Kashimura J, Nagai Y. Inhibitory effect of palatinose on glucose absorption in everted rat gut. *J Nutr Sci Vitaminol (Tokyo).* 2007;53(1):87-9.
- [113] Dietrich CG, Geier A, Oude Elferink RP. ABC of oral bioavailability: transporters as gatekeepers in the gut. *Gut.* 2003;52(12):1788-95.
- [114] Barthe L, Woodley JF, Kenworthy S, Houin G. An improved everted gut sac as a simple and accurate technique to measure paracellular transport across the small intestine. *Eur J Drug Metab Pharmacokinet.* 1998;23(2):313-23.
- [115] Wilson TH, Wiseman G. The use of sacs of everted small intestine for the study of the transference of substances from the mucosal to the serosal surface. *J Physiol.* 1954;123(1):116-25.
- [116] Barthe L, Woodley J, Houin G. Gastrointestinal absorption of drugs: methods and studies. *Fundam Clin Pharmacol.* 1999;13(2):154-68.
- [117] MacLean C, Moenning U, Reichel A, Fricker G. Closing the gaps: a full scan of the intestinal expression of p-glycoprotein, breast cancer resistance protein, and multidrug resistance-associated protein 2 in male and female rats. *Drug Metab Dispos.* 2008;36(7):1249-54.
- [118] Merino G, Jonker JW, Wagenaar E, van Herwaarden AE, Schinkel AH. The breast cancer resistance protein (BCRP/ABCG2) affects pharmacokinetics, hepatobiliary excretion, and milk secretion of the antibiotic nitrofurantoin. *Mol Pharmacol.* 2005;67(5):1758-64.
- [119] Allen JD, van Loevezijn A, Lakhai JM, van der Valk M, van Tellingen O, Reid G, et al. Potent and specific inhibition of the breast cancer resistance protein multidrug transporter in vitro and in mouse intestine by a novel analogue of fumitremorgin C. *Mol Cancer Ther.* 2002;1(6):417-25.



Capítulo V

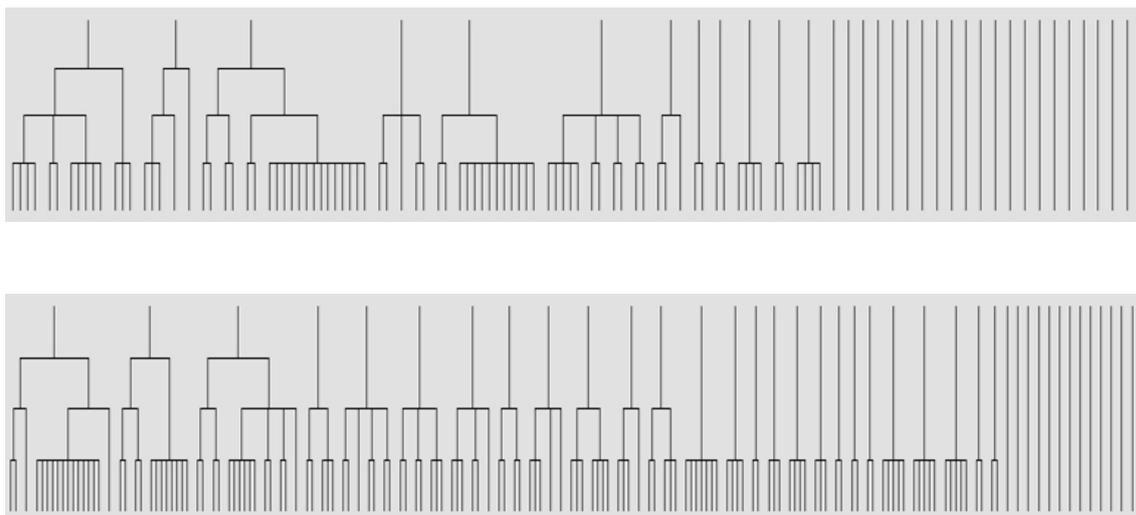
# Resultados

---



## ***5.1 Partición del conjunto de datos en conjunto de entrenamiento y de prueba***

El procedimiento de agrupamiento jerárquico MCS reveló 18 grupos de por lo menos 4 compuestos en la categoría sustratos y 8 grupos en la categoría no sustratos. En la categoría no sustratos hay 1 grupo de 3 compuestos, 6 grupos de 2 compuestos y 22 compuestos que no presentan una subestructura común de 9 átomos con ninguna de las estructuras restantes de esa categoría apareciendo como casos aislados en el dendrograma correspondiente (Figura 5.1 arriba). La categoría sustratos presentó 18 grupos de al menos cuatro compuestos cada uno, 6 grupos de 3 compuestos, 12 grupos de 2 compuestos y 14 compuestos aislados (Figura 5.1 abajo). Es necesario destacar aquí que la sobrerrepresentación de los sustratos en el conjunto de datos (156 sustratos frente a 106 no sustratos) podría estar ligada a la tendencia general de publicar en la literatura científica resultados positivos (en este caso, compuestos capaces de interactuar efectivamente con la BCRP). Mantener esta sobrerrepresentación de los sustratos en el conjunto de entrenamiento podría conducir a modelos sesgados hacia la identificación de sustratos. Para evitar el desbalance en el conjunto de entrenamiento, el 50% de cada grupo definido por el k-means de la categoría de los sustratos y el 75% de cada grupo en la categoría no sustratos fueron asignados al conjunto de entrenamiento, mientras que el porcentaje restante de los grupos se mantuvo como conjunto de prueba independiente. Este esquema permitió la obtención de un conjunto de entrenamiento balanceado en el que ni los sustratos ni los no sustratos se encuentren representados excesivamente, quedando el conjunto de entrenamiento compuesto por 164 compuestos (85 sustratos y 79 no sustratos) y el conjunto de prueba por 98 compuestos (71 sustratos y 27 no sustratos).



**Figura 5.1.** Dendrogramas de los no sustratos (arriba) y sustratos (abajo).

## 5.2 Modelos lineales

### 5.2.1 Primera serie de modelos lineales

Luego de explorar los 102 grupos de descriptores mediante ALD utilizando el método de pasos hacia adelante para la selección de variables, variando la tolerancia entre 0.1 y 0.5, se obtuvieron 196 modelos lineales. Los cuatro mejores modelos individuales que mostraron mayor AUC ROC en el conjunto de prueba están definidos por las siguientes funciones discriminantes:

Modelo 1:

$$Df = -0.859 + 0.011 mlogP^2 - 0.242 nCrR2 + 34.575 JGI7 - 0.424 nCONHR \\ + 0.109 nHAcc - 0.742 GGI8$$

donde  $mlogP^2$  es el cuadrado del coeficiente de partición octanol-agua de Moriguchi,  $nCrR2$  es el número de carbonos (sp<sup>3</sup>) cuaternarios de anillos,  $JGI7$  es la media del índice topológico de carga de orden 7,  $nCONHR$  es el número de

amidas secundarias (alifáticas),  $nHAcc$  es el número de átomos aceptores de enlaces de Hidrógeno (N, O, F) y  $GGI8$  es el índice topológico de carga de orden 8.

Modelo 2:

$$Df = -0.711 + 2.18 BEHm2 - 2.02 BELe2 + 0.26 Hy - 0.74 LAI - 3.20 LP1 \\ + 2.52 BEHp1 - 0.07 SEigp - 1.6E^{-4} VRA2$$

donde  $BEHm2$  es el autovalor n. 2 más alto de la matriz de carga/ponderado por las masas atómicas,  $BELe2$  es el autovalor n. 2 más bajo de la matriz de carga/ponderado por las electronegatividades atómicas de Sanderson,  $Hy$  es el factor hidrofílico,  $LAI$  es el índice de alerta de Lipinski,  $LP1$  es el índice de Lovasz-Pelikan,  $BEHp1$  es el autovalor n. 1 más alto de la matriz de carga/ponderado por las polarizabilidades atómicas,  $SEigp$  es la suma de los autovalores de la matriz de distancia ponderada por la polarizabilidad y  $VRA2$  es el promedio del índice basado en el autovalor de tipo Randic de la matriz de adyacencia.

Modelo 3:

$$Df = 9.80 - 0.006 D/Dr11 - 0.44nCONHR - 0.27 nCO - 10.64 X0Av \\ - 0.14 nCaH - 30.11 Xt - 15.07 PW4 + 0.005 D/Dr12 \\ - 0.001 T(O..O) + 0.38 nNHRPh - 1.7E^{-6} SPI + 0.12 Rww$$

donde  $D/Dr11$  es el índice de distancia/desvío de anillo de orden 11,  $nCONHR$  es el número de amidas secundarias (alifáticas),  $nCO$  es el número de cetonas (alifáticas),  $X0Av$  índice de conectividad de valencia media chi-0,  $nCaH$  es el número de carbonos (sp<sup>2</sup>) aromáticos no sustituidos,  $Xt$  es el índice de conectividad de la estructura total,  $PW4$  es el camino/trayecto 4 - índice de forma Randic,  $D/Dr12$  es el índice de distancia/desvío de anillo de orden 12,  $T(O..O)$  es la suma de las distancias topológicas O..O,  $nNHRPh$  es el número de aminas secundarias (aromáticas),  $SPI$  es el índice super-pendentico y  $Rww$  es el índice de hiperdesvío recíproco.

Modelo 4:

$$Df = -0.64 + 0.009 \text{ mlop}P^2 + 47.43 \text{ JGI7} - 3E^{-5} \text{ SRW10} + 0.01 \text{ piPCO2} \\ + 0.18 \text{ Hy}$$

donde  $\text{mlop}P^2$  es el cuadrado del coeficiente de partición octanol-agua de Moriguchi,  $\text{JGI7}$  es la media del índice topológico de carga de orden 7,  $\text{SRW10}$  es el recuento de trayectos de autorregreso de orden 10,  $\text{piPCO2}$  es el recuento de caminos moleculares múltiples de orden 2 y  $\text{Hy}$  es el factor hidrofílico.

Todos los modelos presentan una excelente relación de casos por cada descriptor incorporado (de 13.7 a 32.8) lo que indica una muy baja probabilidad de sobreajuste. Seguidamente se aplicaron dos métodos de fusión de datos, el operador MAX y el promedio, con lo cual se generaron exhaustivamente todas las posibles combinaciones (o ensambles) de dos de estos 196 modelos. De todas las combinaciones obtenidas, las dos mejores combinaciones resultaron del promedio de los valores del modelo 1 con el 2 (ensamble 1) y del promedio del modelo 3 con el 4 (ensamble 2). En la Tabla 5.1 se presentan los parámetros estadísticos, los porcentajes de buenas clasificaciones y el AUC ROC para el conjunto de entrenamiento y el conjunto de prueba de los cuatro mejores modelos individuales que conforman las 2 mejores combinaciones. Respecto a la validación externa sobre el conjunto de prueba de 98 compuestos, es usual que el porcentaje de aciertos sea mayor en el conjunto de entrenamiento que en un conjunto de prueba independiente, por lo que cabe destacar que en el caso del modelo 4 el porcentaje de clasificaciones correctas global para el conjunto de prueba es mayor respecto al obtenido para el conjunto de entrenamiento. En relación al AUC ROC del conjunto de entrenamiento el modelo 3 supera estadísticamente a los otros tres modelos individuales ( $p = 0.0013$ ,  $0.0037$  y  $< 0.0001$  con respecto al modelo 1, 2 y 4 respectivamente), pero cuando consideramos el AUC ROC del conjunto de prueba, no se encontraron diferencias significativas entre ninguno de los 4 modelos ( $p > 0.05$  en todos los casos). El modelo 3 presenta el mayor número de descriptores de los cuatro modelos y es el que muestra la mayor caída en la capacidad predictiva entre el conjunto de entrenamiento y el de prueba (es decir, una mayor tendencia al sobreajuste).

---

Como era de esperar, el poder explicativo de los modelos obtenidos por aleatorización (ver Tabla 5.2) es significativamente inferior al del modelo real (no aleatorio), ya que la correlación entre la estructura molecular y la propiedad modelada es abolida cuando se aleatoriza la variable dependiente (en nuestro caso, la etiqueta de clase) entre los compuestos del conjunto de entrenamiento. Por otro lado, el desempeño para el caso de la validación LGO, si bien se mantiene cercano, cae respecto al de los modelos originales. Aquí cabe destacar que, debido a la gran diversidad estructural contenida en la base de datos, estamos en presencia de una gran cantidad de compuestos atípicos (como se observó en la etapa de agrupamiento), por lo que en muchos casos se verificó que al ser removidos del conjunto de entrenamiento, varios de estos compuestos son mal clasificados.

**Tabla 5.1.** Características de los cuatro mejores modelos individuales que componen las dos mejores combinaciones seleccionadas.

Modelos	F	Valor de p	Conjunto de entrenamiento				Conjunto de prueba			
			<i>Sp</i> *	<i>Se</i> *	% BC global*	AUC ROC	<i>Sp</i> *	<i>Se</i> *	% BC global*	AUC ROC
Modelo 1	8.04	<.000000	79%	68%	74%	0.796	63%	74%	66%	0.748
Modelo 2	7.52	<.000000	75.3%	74.7%	75%	0.813	76%	66.7%	73.5%	0.750
Modelo 3	10.39	<.000000	83.5%	83.5%	83.5%	0.912	73.2%	74%	73.5%	0.769
Modelo 4	6.56	<.000014	63.3%	70.6%	67%	0.740	77.5%	70.4%	75.5%	0.749

\* considerando cero como valor de corte entre los sustratos y los no sustratos. Este umbral puede optimizarse posteriormente a través de análisis de las curvas ROC para proporcionar un adecuado balance entre *Sp* y *Se*. La *Sp* representa el porcentaje de buenas clasificaciones de los sustratos y la *Se* el porcentaje de buenas clasificaciones de los no sustratos. %BC: porcentajes de buenas clasificaciones.

**Tabla 5.2.** Resultados de la validación interna de los cuatro mejores modelos individuales que componen las dos mejores combinaciones seleccionadas.

Modelos	LGO <sup>1</sup>	Aleatorización <sup>2</sup>
Modelo 1	70.4% (11.9)	64.4% (3.4)
Modelo 2	67% (15)	61.5% (3.6)
Modelo 3	81.2% (11.3)	62.4% (5.1)
Modelo 4	64.8% (13.6)	58.3% (4.05)

<sup>1</sup>Los resultados se presentan como el resultado promedio de las 50 repeticiones y la desviación estándar entre paréntesis. <sup>2</sup> Los resultados se presentan como el rendimiento promedio de los 50 modelos aleatorizados y la desviación estándar entre paréntesis. Se consideró el cero como valor de corte entre los sustratos y los no sustratos.

En la Tabla 5.3 se presentan los resultados de los cuatro modelos individuales aplicados a las bibliotecas de mayor tamaño generadas especialmente para desafiar los modelos durante la validación externa. En la Tabla 5.4 se muestra el comportamiento de las dos mejores combinaciones en las cuatro bases de datos evaluadas.

**Tabla 5.3.** Valores de *Se* y *Sp*, porcentaje de buenas clasificaciones (%BC) global y AUC ROC de los cuatro mejores modelos individuales para la biblioteca simulada de 577 compuestos y la biblioteca de señuelos DUD-E de 1346 compuestos.

	Biblioteca simulada				Biblioteca de señuelos DUD-E			
	<i>Sp</i>	<i>Se</i>	% BC global	AUC ROC	<i>Sp</i>	<i>Se</i>	% BC global	AUC ROC
<b>Modelo 1</b>	66	74	66	0.732	46	74.1	46.6	0.622
<b>Modelo 2</b>	61.1	66.7	61.3	0.689	57.2	66.7	57.4	0.662
<b>Modelo 3</b>	71.8	74.1	72	0.761	49	74.1	49.5	0.612
<b>Modelo 4</b>	63	70.4	63.3	0.681	46	70.4	46.5	0.602

Se consideró el cero como valor de corte entre los sustratos y los no sustratos.

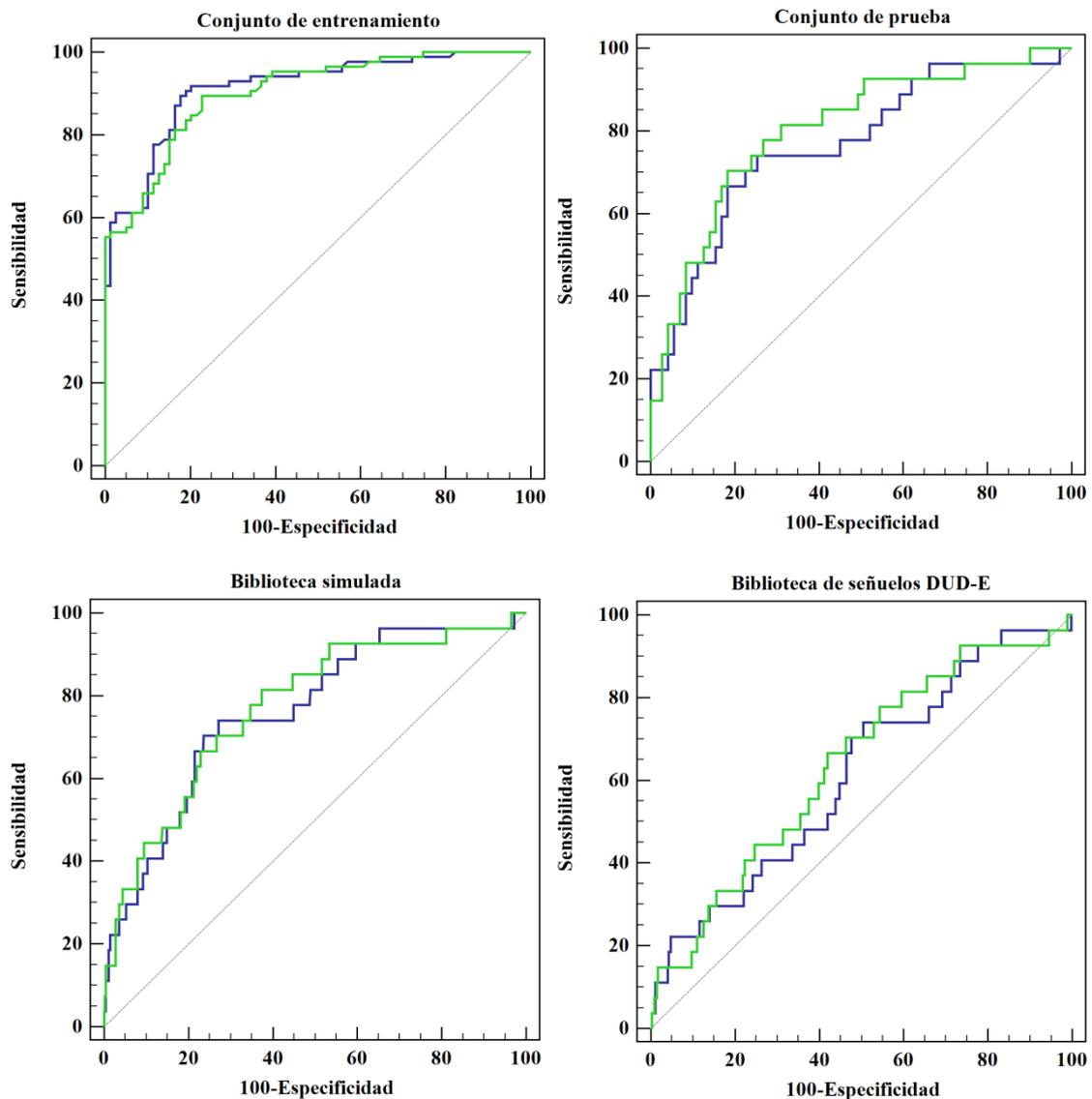
**Tabla 5.4.** Valores de *Se* y *Sp*, porcentaje de buenas clasificaciones (%BC) global y AUC ROC de las dos mejores combinaciones tanto para el conjunto de entrenamiento, el conjunto de prueba, la biblioteca simulada y la biblioteca de señuelos DUD-E.

	Ensamble 1				Ensamble 2			
	<i>Sp</i>	<i>Se</i>	% BC global	AUC ROC	<i>Sp</i>	<i>Se</i>	% BC global	AUC ROC
<b>Conjunto de entrenamiento</b>	84	75	79	0.850	85	80	82	0.902
<b>Conjunto de prueba</b>	70	74	71	0.785	76	70	74.5	0.804
<b>Biblioteca simulada</b>	64	74	64	0.736	68	70	68	0.771
<b>Biblioteca de señuelos DUD-E</b>	53.6	74.1	54.1	0.660	48	70.4	48.5	0.637

Se consideró el cero como valor de corte entre los sustratos y los no sustratos.

La comparación estadística del AUC ROC del conjunto de entrenamiento demostró que el ensamble 2 es estadísticamente superior al ensamble 1 ( $p = 0.0413$ ), mientras que con respecto al modelo 3 (que supera los otros tres modelos individuales) no hay diferencias estadísticas ( $p = 0.4415$ ). Al evaluar el conjunto de prueba, no se encontraron diferencias estadísticamente significativas en el AUC ROC entre ninguno de los modelos individuales y combinaciones ( $p > 0.05$ ). Lo mismo sucede cuando evaluamos el AUC ROC de la biblioteca simulada y la biblioteca de señuelos DUD-E. Observando el AUC ROC en las Tablas 5.1, 5.3 y 5.4, se puede apreciar una caída en el comportamiento de los modelos y combinaciones a medida que va aumentando el tamaño y la diversidad del conjunto de datos analizado, siendo siempre mejor en el conjunto de entrenamiento y cayendo hacia el conjunto de prueba, la librería simulada y la librería de señuelos DUD-E. Esta caída se hace aún más pronunciada en esta última, la cual constituye nuestra base de datos más desafiante y que más se aproxima a una situación de cribado virtual real. Este resultado indica la necesidad de continuar con la búsqueda de modelos que presenten un mejor desempeño de clasificación, en particular, en la biblioteca de señuelos DUD-E de 1346 compuestos. En la Figura 5.2 se muestran las curvas ROC del modelo 3 y del ensamble 2 para las cuatro bases de datos, donde se puede

apreciar la caída en la métrica de enriquecimiento de los modelos a medida que se van aplicando a bases de datos de mayor tamaño y diversidad durante la validación externa.



**Figura 5.2.** Curvas ROC del conjunto de entrenamiento, conjunto de prueba, biblioteca simulada y biblioteca de señuelos DUD-E para el modelo 3 (azul) y el ensamble 2 (verde). Se observa un comportamiento similar del modelo individual y de la combinación y se puede apreciar la caída en la métrica de enriquecimiento de ambos cuando son aplicados a bases de datos de mayor tamaño y diversidad durante la validación externa.

### 5.2.2 Segunda serie de modelos lineales

A través de la aplicación del ERM se obtuvieron 459 modelos que incorporan entre 4 y 12 descriptores moleculares. 97 modelos mostraron un porcentaje global de buenas clasificaciones por encima del 70% en el conjunto de prueba independiente. Todos ellos presentaron buenos resultados en la validación cruzada LOO. En la Tabla 5.5 se muestran las características de los 10 modelos que mostraron la mayor AUC ROC para el conjunto de prueba. Todos los modelos individuales mostraron un AUC ROC estadísticamente diferente del azar (clasificación aleatoria AUC ROC = 0.5) por debajo de un nivel de 0.0001 para el conjunto de entrenamiento, conjunto de prueba y la biblioteca simulada, por lo que cabe destacar que el desempeño del ERM para la selección de variables resultó superior al desempeño del método de pasos hacia adelante utilizado en la etapa anterior. Sin embargo, nuevamente se observó una caída en el AUC ROC de los modelos a medida que fueron aplicados a las bibliotecas de mayor tamaño y diversidad, siendo esta caída abrupta en la librería de señuelos DUD-E.

**Tabla 5.5.** Características de los 10 mejores modelos individuales obtenidos con el ERM. Se toma como valor de corte entre las clases un valor de score igual a cero para evaluar la *Sp* y la *Se*. %BC: porcentajes de buenas clasificaciones.

Modelo	F	Conjunto de entrenamiento				Conjunto de prueba				Biblioteca simulada				Biblioteca de señuelos DUD-E			
		<i>Sp</i>	<i>Se</i>	% BC global	AUC ROC	<i>Sp</i>	<i>Se</i>	% BC global	AUC ROC	<i>Sp</i>	<i>Se</i>	% BC global	AUC ROC	<i>Sp</i>	<i>Se</i>	% BC global	AUC ROC
M438	17.588	91	81	86	0.945	85	52	75.5	0.823	76	52	74.5	0.755	51.8	61.6	61.4	0.503
M380	17.415	86	85	85	0.939	72	78	73.5	0.789	68	78	67	0.746	77.8	57.5	57.9	0.576
M107	18.373	82	85	83.5	0.894	73	67	71	0.788	61	67	61	0.719	66.7	48.6	49	0.636
M471	18.666	86	90	88	0.957	76	70	74.5	0.788	75	70	74.5	0.765	70.4	45.4	45.9	0.611
M356	17.841	87	84	85	0.946	79	59	73.5	0.783	70	59	69.5	0.747	59.3	51	51.2	0.591
M61	19.211	86	71	79	0.885	75	63	71	0.774	70	63	67	0.756	63	46.7	47	0.666
M203	18.426	82	84	83	0.924	70.4	70.4	70	0.774	67	70.4	71	0.736	70.4	57.4	57.6	0.590
M289	17.838	85	82	83.5	0.936	59	80	74.5	0.774	71.5	80	69.5	0.751	59.3	50.1	50.3	0.580
M36	19.970	88	78	83.5	0.863	74	76	75.5	0.773	65	76	66	0.725	74.1	47.4	47.9	0.634
M446	17.419	86	85	85	0.948	77.5	55.5	71	0.772	66	55.5	65	0.732	55.5	51.1	51.2	0.610

Se consideró el cero como valor de corte entre los sustratos y los no sustratos.

La fuerte caída en todos los modelos individuales cuando fueron aplicados a la librería de señuelos DUD-E demuestra nuevamente la dificultad de encontrar una única relación lineal capaz de clasificar con precisión tanto a los sustratos como a los no sustratos cuando nos enfrentamos a una aplicación de cribado virtual real. En este sentido, hay que tener en cuenta (como ya se mencionó en el Capítulo IV) que la poliespecificidad de los transportadores ABC debida a la presencia de múltiples sitios de unión y a su alta flexibilidad, determina un fenómeno complejo que hace a este problema de modelado particularmente difícil. Por este motivo se combinaron por un lado, los 10 modelos con la mayor AUC ROC en el conjunto de prueba y por otro lado, los 97 modelos con un porcentaje global de buena clasificaciones por encima del 70% en el conjunto de prueba a través de cinco esquemas de fusión de datos: el operador MAX, el operador MIN, el promedio, el ranking promedio y el voto promedio (Tabla 5.6).

**Tabla 5.6.** Comparación del AUC ROC de las combinaciones de los 10 y 97 mejores modelos obtenidos por ERM a través de los cinco esquemas de fusión de datos para el conjunto de entrenamiento, el conjunto de prueba, la biblioteca simulada y la biblioteca de señuelos DUD-E.

	AUC ROC			
	Conjunto de entrenamiento	Conjunto de prueba	Biblioteca simulada	Biblioteca de señuelos DUD-E
<b>Combinación de 10 modelos</b>				
Operador MAX	0.954	0.850	0.824	0.743
Operador MIN	0.939	0.776	0.724	0.526
Promedio	0.963	0.833	0.791	0.628
Ranking promedio	0.963	0.830	0.789	0.629
Voto promedio	0.871	0.724	0.711	0.558
<b>Combinación de 97 modelos</b>				
Operador MAX	0.829	0.647	0.530	0.589
Operador MIN	0.934	0.740	0.573	0.512
Promedio	0.964	0.778	0.74	0.598

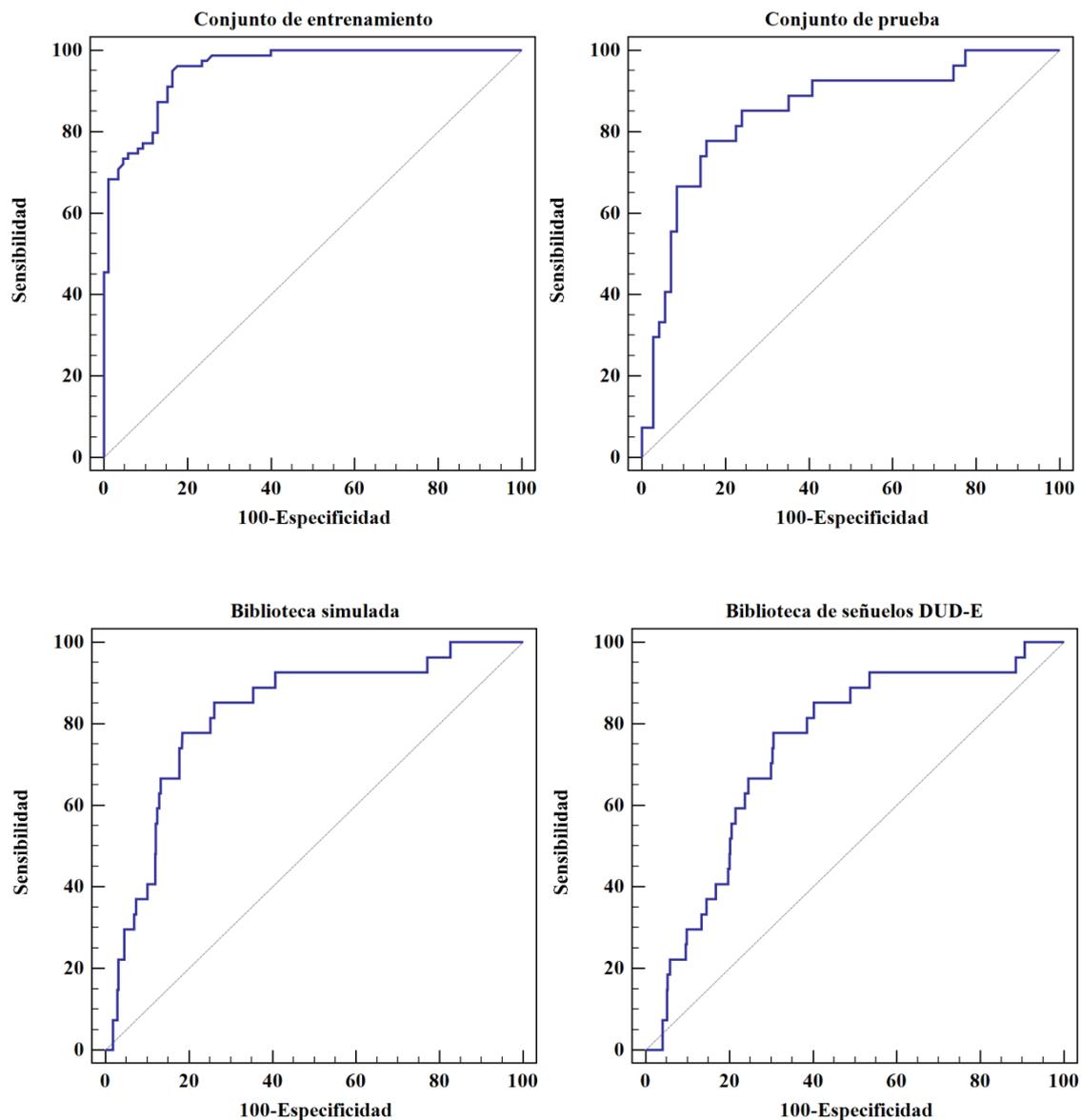
Ranking promedio	0.963	0.781	0.745	0.591
Voto promedio	0.906	0.712	0.651	0.622

La combinación de los 97 modelos mediante los cinco esquemas de fusión de datos, mostró la misma caída abrupta en la capacidad predictiva que se observó en los modelos individuales durante la validación externa. Sin embargo, la combinación de los 10 mejores modelos mediante el operador MAX dio los mejores resultados, con un AUC ROC de 0.824 en la biblioteca simulada y de 0.743 en la biblioteca de señuelos DUD-E, siendo el único esquema de fusión de datos que demostró un buen rendimiento en la última y mayor biblioteca evaluada; mientras que para esta biblioteca, todos los otros modelos y combinaciones mostraron o bien muy pobre o ningún poder clasificatorio (ver Tabla 5.6). Esto está de acuerdo con la noción de ensamble o combinación selectiva de clasificadores que se introdujo en el Capítulo IV, según la cual la combinación de unos pocos y selectos clasificadores puede proporcionar predicciones más robustas que la combinación de todos los modelos disponibles. De acuerdo a esto, también se evaluaron las combinaciones de los 2, 3, 5, y 15 modelos con la mejor AUC ROC en el conjunto de prueba. A pesar de que ninguna de estas combinaciones mostró diferencias estadísticamente significativas (a un nivel de 0.05) en el AUC ROC para el conjunto de entrenamiento, el conjunto de prueba y la biblioteca simulada; el comportamiento observado para la biblioteca de señuelos DUD-E demuestra que la mejor combinación es la de 10 modelos, que superó a las combinaciones de 2, 3 y 5 modelos con un nivel de significancia estadística de  $p < 0.0001$ . Mientras que la combinación de 15 modelos mostró un rendimiento muy similar a la combinación de 10 modelos (AUC ROC de 0.743 para la combinación de 10 modelos frente a un AUC ROC de 0.738 para la de 15 modelos, sin diferencia estadísticamente significativa  $p = 0.6679$ ), resultando esta última la combinación óptima.

Nuevamente, tanto para los modelos individuales como para las combinaciones (sin importar el esquema de fusión de datos utilizado) el rendimiento en el conjunto de prueba de 98 compuestos fue generalmente mejor que el rendimiento en la biblioteca simulada de 577 compuestos, y cae en la biblioteca de señuelos DUD-E. Este resultado parece confirmar que la evaluación

del desempeño de los modelos sobre una base de datos de mayor tamaño y diversidad estructural es una prueba más rigurosa que la evaluación del desempeño en una base de datos pequeña, y demuestra la capacidad de la biblioteca de señuelos DUD-E para generar señuelos adecuados con fines de validación.

En la Figura 5.3 se muestran las curvas ROC de la mejor combinación obtenida para las cuatro bases de datos evaluadas.



**Figura 5.3.** Curvas ROC de la mejor combinación obtenida. La combinación de 10 modelos a través del operador MAX es el único enfoque que mantuvo el poder clasificadorio en todas las bibliotecas utilizadas para la validación externa.

Si bien se obtuvo una combinación que mantuvo su poder predictivo en todas las bases de datos utilizadas durante la validación externa, se decidió aumentar la complejidad y realizar un nuevo y último proceso de modelado durante el cual se decidió recurrir a modelos de tipo no lineal, para lo cual se utilizó una combinación de algoritmos genéticos (GA) con el algoritmo de inducción de árboles de decisión J48.

### 5.3 Modelos No Lineales

Mediante la aplicación de GA se obtuvieron 19 subconjuntos de entre 5 a 20 descriptores para cada uno de los 6 conjuntos de descriptores aleatorizados (114 subconjuntos en total), los cuales resultan de mayor utilidad para separar las dos categorías de compuestos consideradas. A partir del análisis exhaustivo de estos 114 subconjuntos mediante la variación de los parámetros de corrida del algoritmo J48 según se detalló en el Capítulo IV, se seleccionaron 12 modelos que presentaban mejor desempeño tanto para el conjunto de prueba como para la biblioteca simulada. Los diagramas de estos 12 árboles de decisión se detallan en el Anexo II. En la Tabla 5.7 se muestran los porcentajes de buenas clasificaciones globales para el conjunto de entrenamiento y de prueba seguidos de los resultados de la validación interna (LGO y aleatorización), mientras que en la Tabla 5.8 se muestran los resultados del AUC ROC del conjunto de entrenamiento, conjunto de prueba, la biblioteca simulada y la biblioteca de señuelos DUD-E para los 12 modelos seleccionados.

**Tabla 5.7.** Porcentajes de buenas clasificaciones (%BC) global del conjunto de entrenamiento y conjunto de prueba y resultados de la validación interna para los 12 mejores modelos obtenidos.

Modelo	% BC global conjunto de entrenamiento	% BC global conjunto de prueba	LGO <sup>1</sup>	Aleatorización <sup>2</sup>
Modelo 1	73.78	73.47	63.93 (13.08)	58.98 (2.33)
Modelo 2	95.73	73.47	68.55 (16.13)	56.75 (5.73)
Modelo 3	93.29	74.49	62.81 (14.56)	56.16 (4.81)
Modelo 4	95.12	69.39	68.37 (14.34)	57.16 (4.51)
Modelo 5	96.95	68.37	69.95 (13.95)	58.16 (4.66)
Modelo 6	96.34	67.34	70.60 (14.14)	58.53 (5.06)
Modelo 7	92.07	76.53	68.22 (12.33)	57.69 (5.76)
Modelo 8	83.54	78.57	68.57 (12.31)	57.38 (5.67)
Modelo 9	78.05	75.51	65.93 (12.97)	54.12 (4.61)

Modelo 10	91.46	75.51	64.19 (13.98)	54.44 (4.59)
Modelo 11	82.32	77.55	66.57 (13.87)	56.64 (4.96)
Modelo 12	91.46	75.51	67.92 (14.28)	56.71 (4.90)

<sup>1</sup> Los resultados se presentan como el resultado promedio de las 150 repeticiones y la desviación estándar entre paréntesis. <sup>2</sup> Los resultados se presentan como el rendimiento promedio de los 50 modelos aleatorizados y la desviación estándar entre paréntesis. Se consideró un valor de corte de 0.5 entre los sustratos y los no sustratos.

**Tabla 5.8.** Valores de AUC ROC del conjunto de entrenamiento, conjunto de prueba, biblioteca simulada y biblioteca de señuelos DUD-E para los 12 mejores modelos obtenidos.

Modelo	AUC ROC conjunto de entrenamiento	AUC ROC conjunto de prueba	AUC ROC biblioteca simulada	AUC ROC biblioteca de señuelos DUD-E
Modelo 1	0.785	0.737	0.696	0.615
Modelo 2	0.992	0.791	0.735	0.675
Modelo 3	0.968	0.829	0.734	0.701
Modelo 4	0.993	0.788	0.718	0.598
Modelo 5	0.996	0.785	0.716	0.590
Modelo 6	0.995	0.788	0.718	0.598
Modelo 7	0.969	0.856	0.780	0.809
Modelo 8	0.867	0.744	0.725	0.742
Modelo 9	0.815	0.880	0.820	0.651
Modelo 10	0.965	0.850	0.758	0.757
Modelo 11	0.843	0.728	0.707	0.681
Modelo 12	0.962	0.839	0.795	0.756

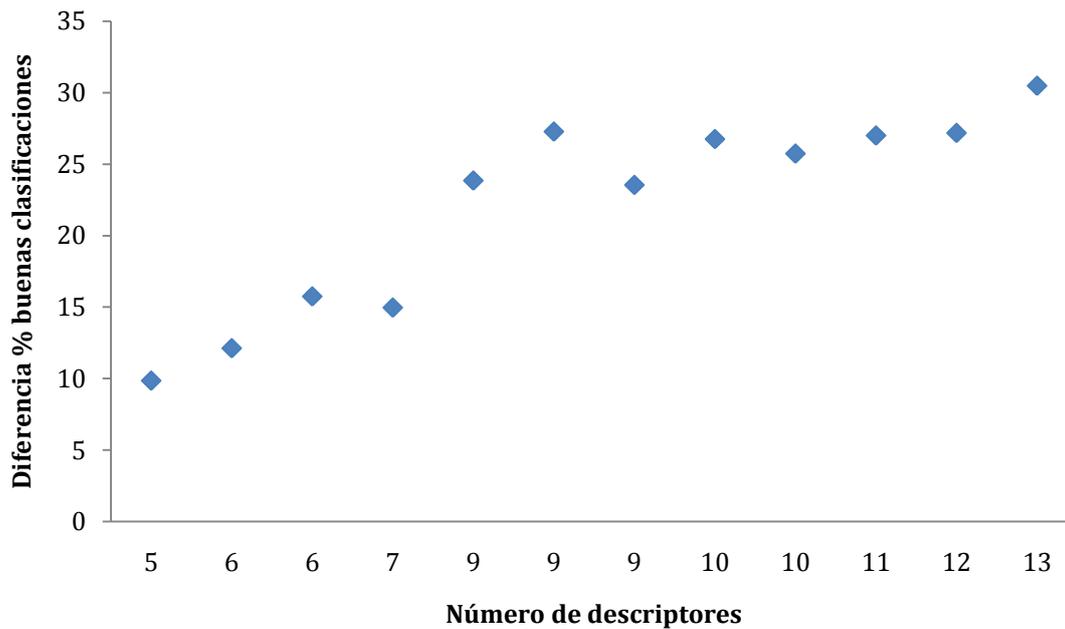
En la Tabla 5.7 puede advertirse que el porcentaje de buenas clasificaciones globales es menor para el conjunto de prueba que para el conjunto de entrenamiento prácticamente para todos los modelos. En la Tabla 5.8 se observa que en la mayoría de los modelos nuevamente hay una tendencia de descenso en la

capacidad de predicción desde el conjunto de prueba hacia la biblioteca de señuelos DUD-E, con algunas excepciones como los modelos 7, 8, 10 y 12 en los cuales el desempeño se mantiene a lo largo de las bases de datos evaluadas. Si bien todos los modelos presentan buenos resultados en la validación interna, esta tendencia de descenso se observa también en los resultados de la validación interna LGO (Tabla 5.7). En particular se vio que al aumentar el número de descriptores incorporados al modelo, aumentó la diferencia entre el porcentaje de buenas clasificaciones del conjunto de entrenamiento y el de las instancias removidas durante el proceso de validación (ver Tabla 5.9 y Figura 5.4). Es notable que el algoritmo J48 presenta una fuerte tendencia al sobreajuste con los datos del conjunto de entrenamiento que se acentúa con el número de variables independientes del modelo, lo cual es un problema general de los métodos altamente flexibles [1-2], si bien durante el proceso de obtención de los modelos se hizo gran hincapié en utilizar las distintas herramientas de poda disponibles que, como ya se analizó en el Capítulo IV, permiten controlar en gran medida esta tendencia.

**Tabla 5.9.** Diferencia entre los porcentajes de buenas clasificaciones (%BC) del conjunto de entrenamiento y el de las instancias removidas en la validación interna LGO en función del número de descriptores incorporados al modelo.

<b>Modelo</b>	<b>Nº de descriptores en el modelo</b>	<b>Relación Nº casos conjunto de entrenamiento/ Nº descriptores del modelo</b>	<b>Diferencia entre % BC conjunto de entrenamiento y la validación LGO</b>
Modelo 1	5	32.8	9.85
Modelo 2	6	27.33	12.12
Modelo 3	6	27.33	15.74
Modelo 4	7	23.42	14.96
Modelo 5	9	18.22	23.85
Modelo 6	9	18.22	27.27
Modelo 7	9	18.22	23.54
Modelo 8	10	16.4	26.75

Modelo 9	<b>10</b>	16.4	25.73
Modelo 10	<b>11</b>	14.91	26.99
Modelo 11	<b>12</b>	13.66	27.18
Modelo 12	<b>13</b>	12.61	30.48



**Figura 5.4.** Evolución lienal de la diferencia entre los porcentajes de buenas clasificaciones del conjunto de entrenamiento y de las instancias removidas en la validación LGO (eje y) vs el número de descriptores incorporados al modelo (eje x). Se observa una clara tendencia al sobreajuste con el aumento del número de descriptores.

Además, el algoritmo J48 presenta cierto grado de inestabilidad, entendiéndose por esto que, una pequeña variación en el conjunto de los datos de entrenamiento puede dar lugar, en ciertas ocasiones, a un árbol de decisión diferente [2-3]. Este hecho se deriva directamente del procedimiento jerárquico divisivo por el cual se construye el árbol de decisión. Es lógico pensar que, cuando ciertos compuestos del conjunto de entrenamiento son removidos, la elección del descriptor por el cual se produce la división binaria de los datos en cada nodo puede ser distinta y a partir de allí, el resto del árbol generado será distinto.

Nuevamente es destacable la gran variabilidad estructural de los compuestos de la base de datos y el número elevado de compuestos atípicos (aislados) en ambas clases.

Debido a esta inestabilidad y la tendencia al sobreajuste característica de este algoritmo, la combinación de modelos por medio de técnicas de fusión de datos aparece como una posible solución [2-3]. Por este motivo, se generaron combinaciones de los 12 mejores modelos seleccionados (ensamblado selectivo) utilizando los cinco esquemas de fusión de datos ya aplicados en la instancia de modelado anterior: el operador MAX, el operador MIN, el promedio, el ranking promedio y el voto promedio según se detalló en el Capítulo IV. Los resultados se muestran en la Tabla 5.10.

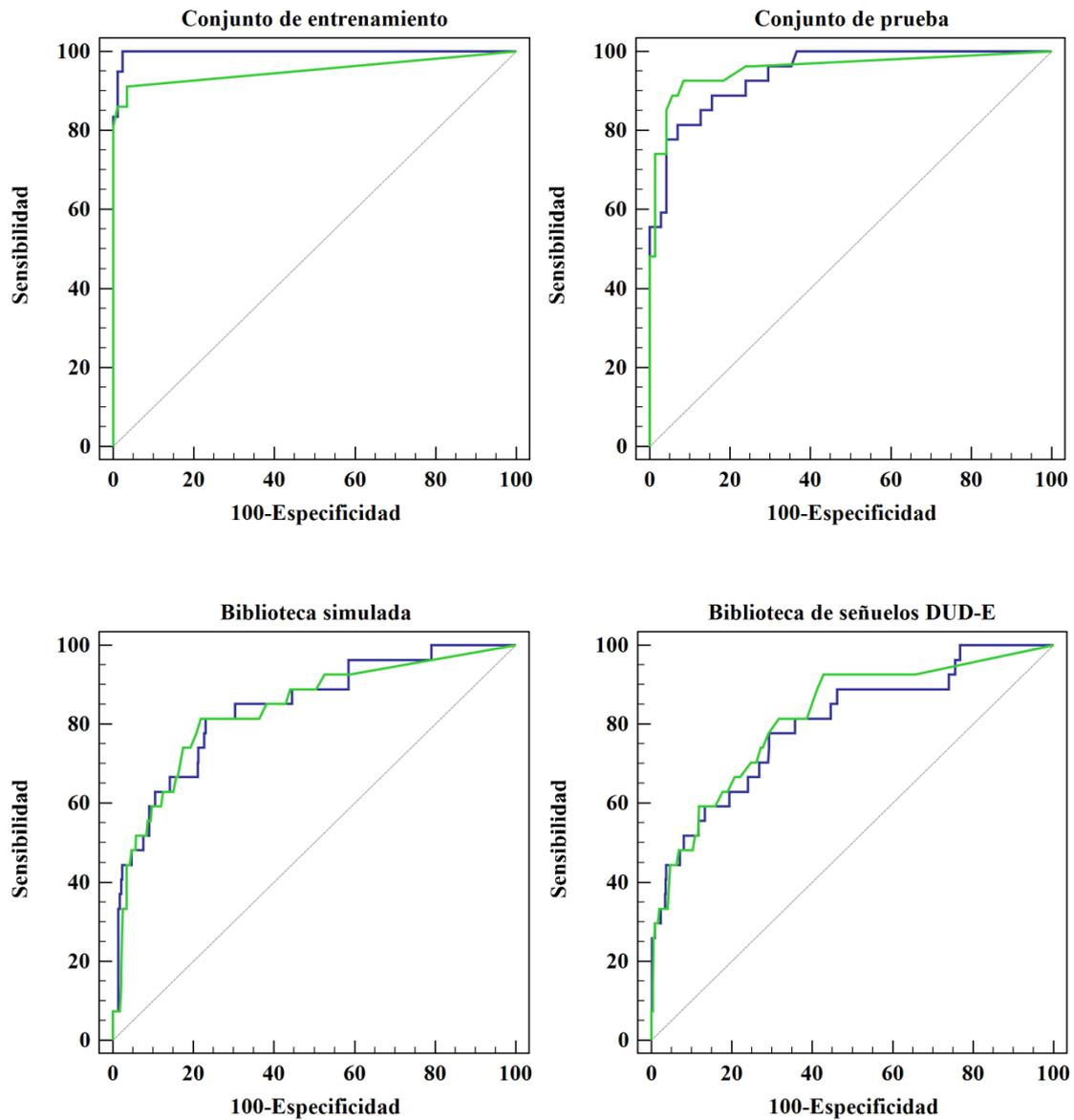
**Tabla 5.10.** Valores de AUC ROC del conjunto de entrenamiento, conjunto de prueba, biblioteca simulada y biblioteca de señuelos DUD-E para los 5 esquemas de fusión de datos aplicados sobre los 12 mejores modelos obtenidos.

Ensamble	AUC ROC conjunto de entrenamiento	AUC ROC conjunto de prueba	AUC ROC biblioteca simulada	AUC ROC biblioteca de señuelos DUD-E
Operador MAX	0.978	0.845	0.741	0.668
Operador MIN	0.986	0.825	0.797	0.786
Promedio	0.994	0.890	0.826	0.779
Ranking Promedio	0.997	0.945	0.840	0.801
Voto promedio	0.952	0.956	0.834	0.818

En este caso, a pesar de mantenerse la biblioteca de señuelos DUD-E como aquella en la que se observa peor desempeño, es la primera vez donde se observa un buen comportamiento para todos los esquemas de fusión de datos (excepto el

operador MAX) a lo largo de todas las bases de datos utilizadas en la validación externa. Respecto al conjunto de entrenamiento, la mayor AUC ROC la presenta el ranking promedio superando al voto promedio con un nivel  $p = 0.0055$ , no habiendo diferencias respecto a los restantes esquemas. En cuanto al conjunto de prueba, las mayores AUC ROC son las del ranking promedio y la del voto promedio no encontrándose diferencias estadísticamente significativas entre ellas ( $p > 0.05$ ), mientras que ambos difieren con los tres restantes en un nivel ( $p < 0.03$ ). Respecto de la biblioteca simulada el valor máximo de AUC ROC es el del ranking promedio, no presentando diferencias con las restantes combinaciones ( $p > 0.05$ ). En la biblioteca de señuelos DUD-E, el mayor valor lo presenta el voto promedio, no teniendo diferencias significativas con los restantes combinaciones, excepto con el operador MAX ( $p < 0.0001$ ).

El mejor comportamiento lo presentan tanto el ranking promedio como el voto promedio de manera consistente en todas las bibliotecas utilizadas en la validación externa, y si bien no se encontraron diferencias estadísticamente significativas entre ellos (excepto en el conjunto de entrenamiento donde resultó superior el ranking promedio), se decidió utilizar el ranking promedio de los 12 mejores árboles de decisión para pasar a la etapa de validación experimental considerando que se trata de un esquema de fusión de datos más sencillo que el voto promedio. En la Figura 5.5 puede observarse el desempeño similar tanto del ranking promedio como del voto promedio mediante las curvas ROC para el conjunto de entrenamiento, el conjunto de prueba, la biblioteca simulada y la biblioteca de señuelos DUD-E. En la Tabla 5.11 se muestra el ordenamiento de los compuestos del conjunto de prueba generado por el promedio del ranking, lo que ilustra tanto el poder clasificador como el poder de enriquecimiento en no sustratos de la primera porción del ordenamiento que presenta este esquema de fusión de datos.



**Figura 5.5.** Curvas ROC para el conjunto de entrenamiento, el conjunto de prueba, la biblioteca simulada y la biblioteca de señuelos DUD-E tanto para el ranking promedio (azul) como para el voto promedio (verde) de los 12 mejores modelos no lineales seleccionados. Se puede observar el desempeño similar de ambos esquemas de fusión de datos. Se utilizó el ranking promedio dado que es un esquema de fusión de datos más sencillo que el voto promedio.

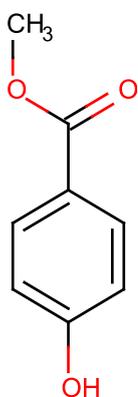
**Tabla 5.11.** Ordenamiento de los compuestos del conjunto de prueba generado por el ranking promedio de los 12 mejores modelos no lineales. En azul se muestran los no sustratos y en rojo los sustratos de la BCRP. Se puede observar el enriquecimiento de no sustratos en la primera porción del ordenamiento.

Molécula	PR	Molécula	PR
NS-SN-349	8	S-SN-392	56
NS-Primidone	8.75	S-Lapatinib (GW572016)	56
NS-SN-351	9	NS-Lamotrigine	57
NS-Vinblastine	9.67	S-Methotrexate diglutama	57
NS-SN-352	10	S-EKI-785	58
NS-Vincristine	10.7	S-IAAP	59
NS-SN-364	11.6	S-Didanosine (ddl)	59
NS-SN-353	12.5	S-Tetracycline	59
NS-Tacrolimus	14.5	S-C-2007	59
NS-Trp-P-1	15.2	S-ST-1968	60
NS-Progesterone	16.8	S-Taurodeoxycholate	60
NS-Tetramethylrosamine	17	S-Estrone-3-sulfate	60
NS-Verapamil	18.3	S-Fluvastatin	62
NS-SN-443	20.9	S-Nitrofurantoin	62
NS-NSC19139	23.3	S-Thioguanine	63
S-Rifampicin	23.4	S-Cholate	64
S-Tamoxifen	23.5	S-Nilotinib	65
NS-SN-444	24.5	S-Methotrexate	65
S-IQ	24.6	S-Fludarabine	66
NS-SN-397	25.3	S-Nifedipine	67
NS-Valproate	27.1	S-Flavopiridol	67
NS-NSC168201	28.7	S-NSC107392	68
NS-m-THPC	28.8	S-ST-1587	68
NS-m-THPP	29.8	S-Gemcitabine	68
S-Plumbagin	31.4	S-NSC73306	68
S-Phenethyl isothiocyanat	31.6	S-Pantoprazole	69
NS-Taurocholate	32	S-Riboflavin	69
S-Pyropheophorbide a me	32.5	S-Troglitazone sulfate	69
S-Vitamin K3	32.8	S-Uric Acid	69
S-PhIP	36	S-PF-407288	69
S-Phytoporphyrin	36.4	S-SN-355	69
NS-Tauroolithocholate sulf	37.3	S-ST-1976	70
S-Imidacrine (Symadex, C	39.3	S-Folic acid	70
S-Omeprazole	43.4	S-Zalcitabine (ddC)	70
NS-NSC375985	43.6	S-SN-38	70
S-NB-506 (Compound A)	43.8	S-Rosuvastatin	71
S-Erlotinib	45.1	S-Norfloxacin	71
S-NX211 (Lurtotecan)	45.8	S-C-1419	71
S-Topotecan	47	S-C-1584	72
S-C-1336	49.2	S-Pravastatin	72
S-Ethidium bromide	49.3	S-UCN-01	73
NS-Leucovorin (Folinic Ac	50.1	S-Ofloxacin	73
S-Nicardipine	50.6	S-Puromycin	74
S-Deoxycholate	51	S-Ulifloxacin	77
S-SN-38 glucuronide	51.1	S-Sulfasalazine	79
S-Erythromycin	51.3	S-Phosphatidylserine	81
NS-NSC120688	53.4	S-Quercetin	83
S-SN-398	53.5	S-Thioguanine riboside	86
S-NSC265473	53.9	S-Tomudex	89

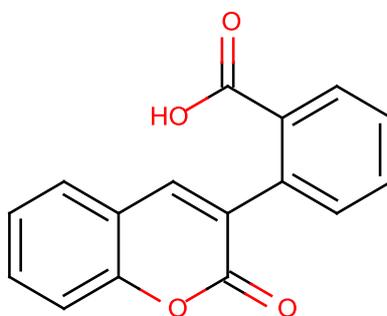
## 5.4 Validación experimental de la combinación seleccionada

### 5.4.1 Selección de los compuestos a evaluar

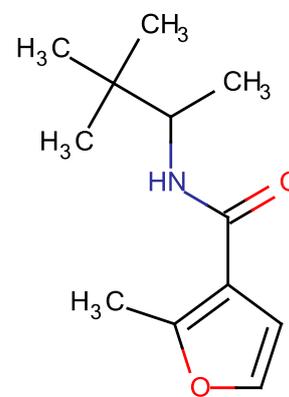
De los 21 compuestos que cumplieron con las condiciones establecidas y detalladas en el Capítulo IV, se seleccionaron los 5 compuestos clasificados como no sustratos que quedaron mejor rankeados en el ordenamiento (es decir, los 5 primeros) utilizando el ranking promedio de los 12 mejores árboles de decisión como tamiz *in silico*. Los 5 compuestos anticonvulsivos seleccionados para su evaluación experimental fueron: metilparabeno (MTP) [4], 2-(2-oxo-2H-chromen-3-yl)benzoic acid (BB\_NC-1569), 1-methyl-1,2,3,4'-tetrahydro-1'H-spiro[indole-3,2'-quinazoline]-2,4'-dione (OSSK\_784104), N-(3,3-dimethylbutan-2-yl)-2-methylfuran-3-carboxamide (OSSL\_216854) [5-6] y N,N'-dibencilsulfamida (DBS) [7]. Las estructuras químicas de los 5 compuestos son las siguientes:



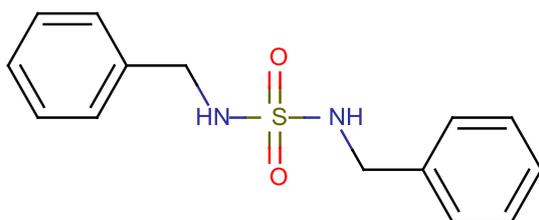
Metilparabeno (MTP)



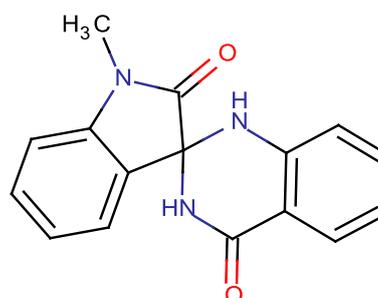
BB\_NC-1569 (BB\_NC)



OSSL\_216854 (OSSL)



N,N'-dibencilsulfamida (DBS)



OSSK\_784104 (OSSK)

La BB\_NC fue adquirida de InterBioScreen Ltd, la OSSK de Princeton Biomolecular Research, la OSSL de UkrOrgSyntez (UORSY) Ltd, el MTP de Sigma-Aldrich Argentina y la DBS fue sintetizada en nuestro laboratorio [7].

### 5.4.2 Condiciones establecidas para la cuantificación por HPLC

En todos los casos la fase móvil estuvo compuesta por:

- A. Buffer  $\text{KH}_2\text{PO}_4$  20 mM, ajustado a pH 2.5 con  $\text{H}_3\text{PO}_4$
- B. Metanol

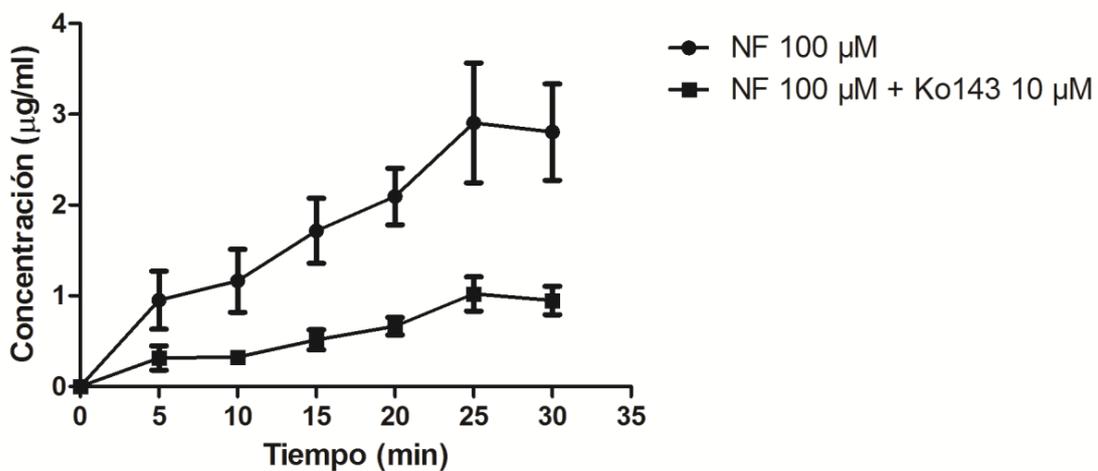
Siguiendo las siguientes condiciones particulares:

	Modo	Relación A:B	Flujo	Longitud de onda de detección
<b>Condiciones para validación del modelo</b>				
NF	Isocrático	60:40	1 ml/min	366 nm
NF + Ko143	Gradiente	0 min: 40% B 2 min: 40% B 4 min: 85% B 6 min: 85% B 7 min: 40% B 10 min: 40% B	1 ml/min	366 nm
<b>Condiciones muestras Protocolo 1</b>				
NF + MTP	Isocrático	55:45	1 ml/min	366 nm
NF + OSSK	Isocrático	50:50	1 ml/min	366 nm
NF + BB_NC	Isocrático	50:50	1.5 ml/min	366 nm
NF + OSSL	Gradiente		1 ml/min	366 nm
NF + DBS	Gradiente		1 ml/min	366 nm
<b>Condiciones muestras Protocolo 2</b>				
MTP	Isocrático	55:45	1 ml/min	254 nm
OSSK	Isocrático	55:45	1 ml/min	224 nm

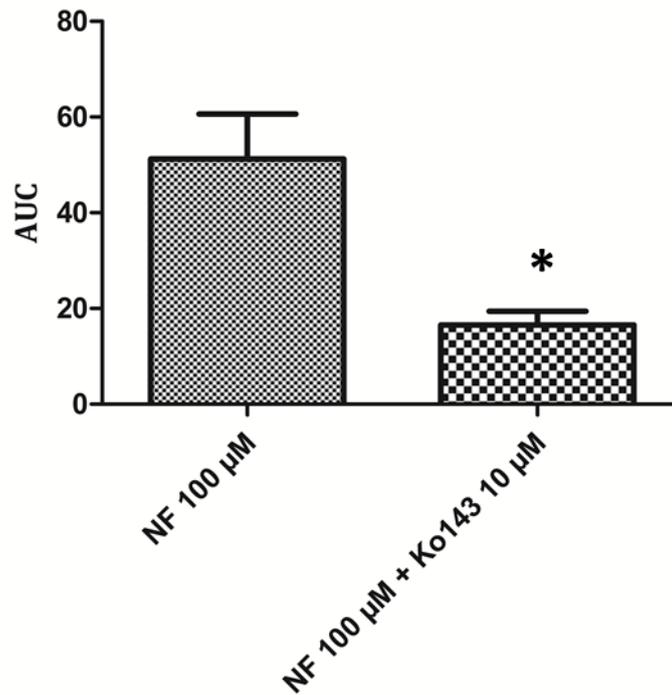
BB_NC	Isocrático	38:62	1 ml/min	300 nm
Drogas + Ko143	Se respetaron las condiciones individuales seguidas de un gradiente para eliminar el Ko143 de la columna			

### 5.4.3 Validación del modelo experimental

Los resultados obtenidos se muestran en las Figuras 5.6 y 5.7, donde se verifica el transporte activo de NF a través del tejido y se observa la inhibición en presencia de Ko143. De todas las concentraciones evaluadas tanto de NF como del inhibidor, la concentración de 100  $\mu\text{M}$  para NF y 10  $\mu\text{M}$  para Ko143 resultaron las concentraciones a las cuales se observó mejor el transporte activo y la inhibición, respectivamente.



**Figura 5.6.** Curvas obtenidas para NF evaluada a 100  $\mu\text{M}$  a 37°C en presencia y ausencia de Ko143 10  $\mu\text{M}$  (n=9). Las barras verticales indican el error estándar de la media (SEM).



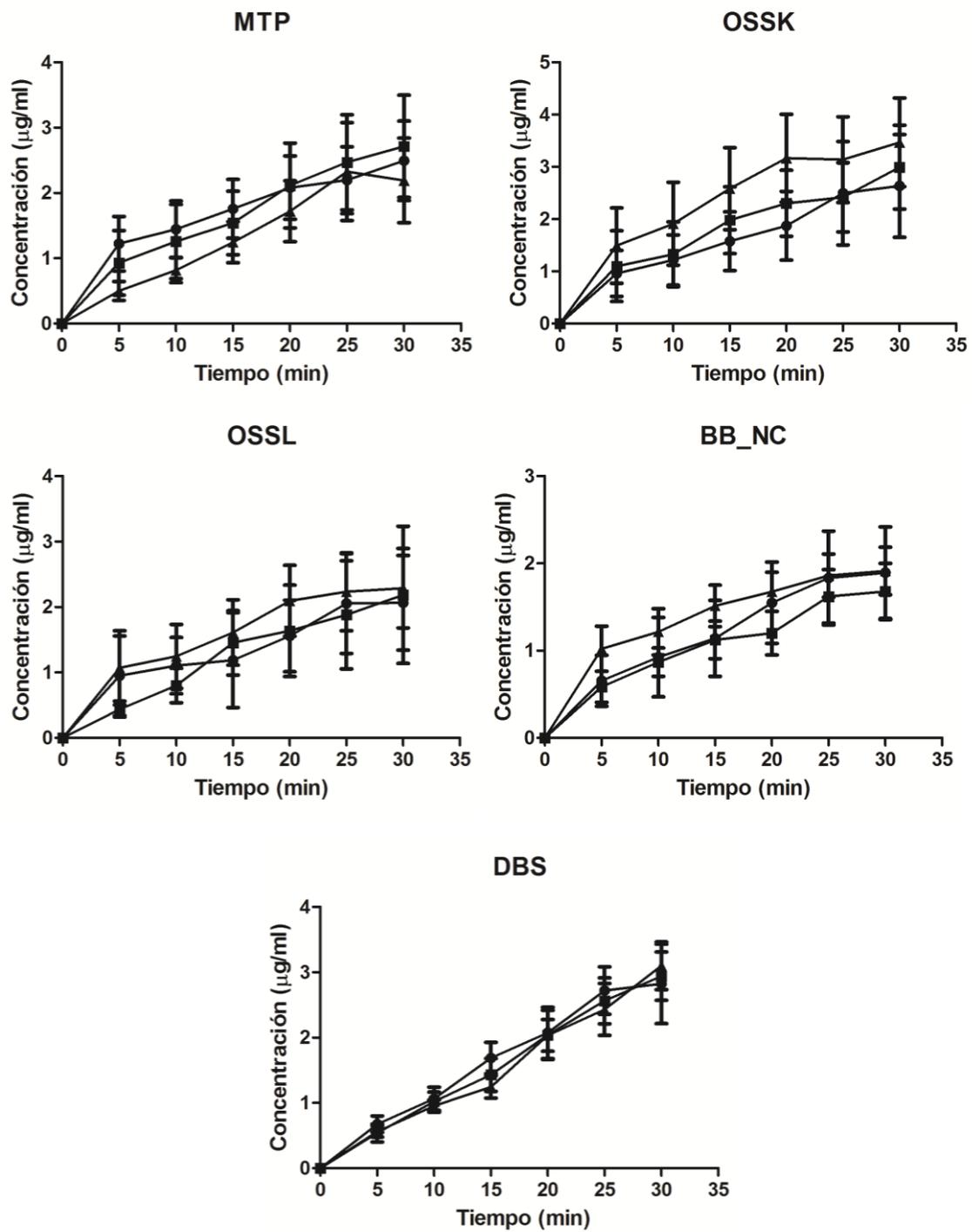
**Figura 5.7.** AUC obtenida para NF evaluada a 100  $\mu\text{M}$  a 37°C en presencia y ausencia de Ko143 10  $\mu\text{M}$  (n=9). Las barras verticales indican el error estándar de la media (SEM).

Se aplicó el test t de Student y se encontraron diferencias estadísticamente significativas entre las AUC de NF sola a 100  $\mu\text{M}$  y en presencia de Ko143 10  $\mu\text{M}$  ( $p = 0.000343$ ). La inhibición mediada por Ko143, al ser un inhibidor específico de la BCRP, permite demostrar que la BCRP efectivamente participa en el transporte activo de NF en las condiciones experimentales utilizadas. Por otro lado, cabe destacar que la mayor dispersión de los datos (barras verticales de las Figuras 5.6 y 5.7) se observó en la curva correspondiente a NF 100  $\mu\text{M}$ , mientras que la dispersión fue mucho menor en presencia del inhibidor. Este comportamiento puede deberse a la variabilidad intraindividual (diferente porción del intestino) así como interindividual del nivel de expresión de BCRP *in vivo*, afectando así el transporte activo de NF, efecto que desaparece en presencia del inhibidor específico. Esta observación resulta interesante dado que en los seres humanos es esperable un comportamiento similar.

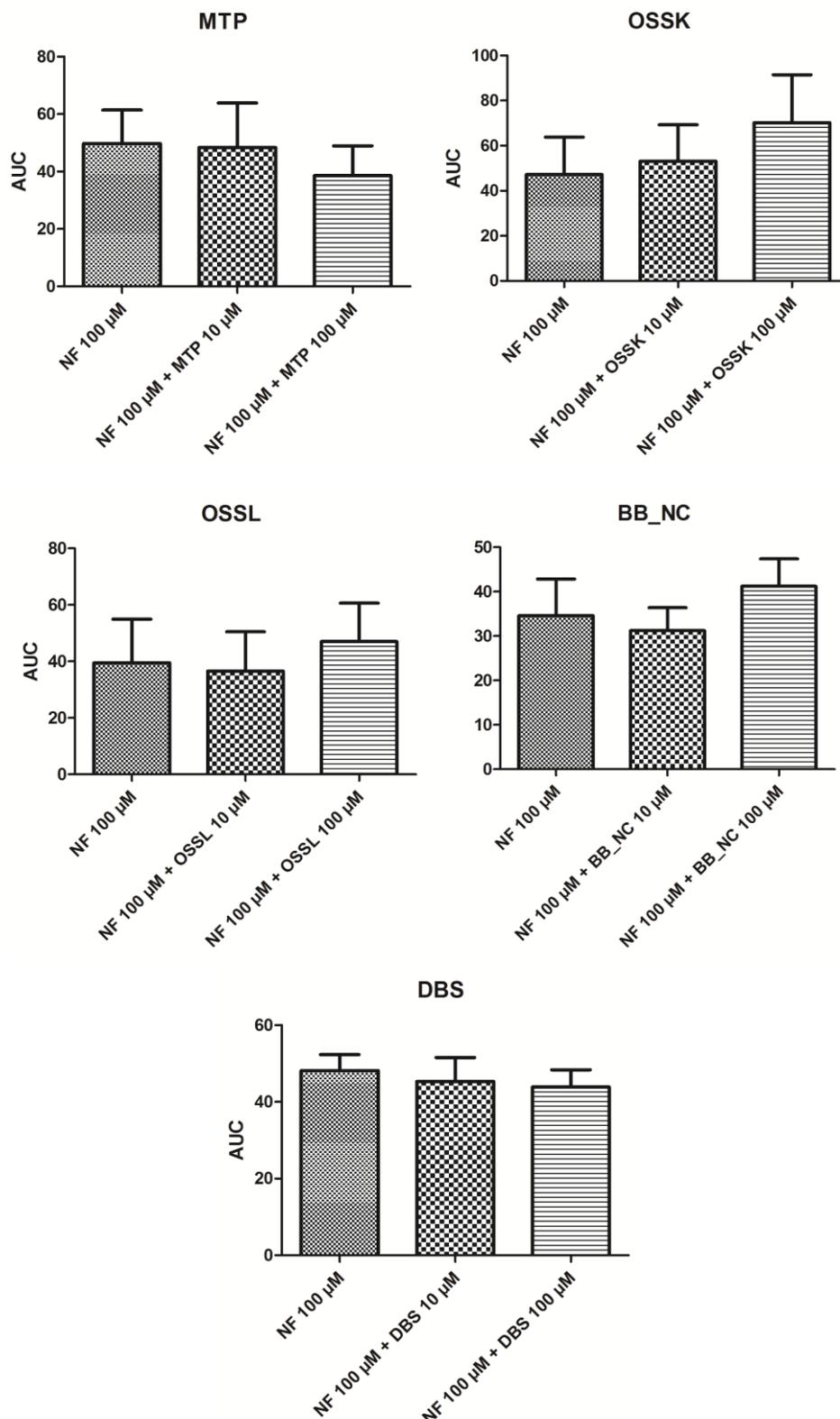
A partir de estos resultados se estableció como concentración óptima de NF 100  $\mu\text{M}$  y como concentración óptima de inhibidor 10  $\mu\text{M}$ .

#### *5.4.4 Evaluación de las drogas según protocolo 1*

En el protocolo 1 se evaluó el transporte activo de NF mientras que las 5 drogas fueron disueltas en el medio exterior a 10 y a 100  $\mu\text{M}$ . Dichas concentraciones, como se detalló en el Capítulo IV, fueron establecidas de acuerdo a las concentraciones óptimas halladas en el punto anterior para el inhibidor (10  $\mu\text{M}$ ) y la NF (100  $\mu\text{M}$ ). Los resultados obtenidos se muestran en las Figuras 5.8 y 5.9.



**Figura 5.8.** Curvas obtenidas para NF evaluada a 100 µM en ausencia (curva control ●) y en presencia de la droga ensayada a 10 (■) y 100 µM (▲) a 37 °C (n=6). Las barras verticales indican el error estándar de la media (SEM).



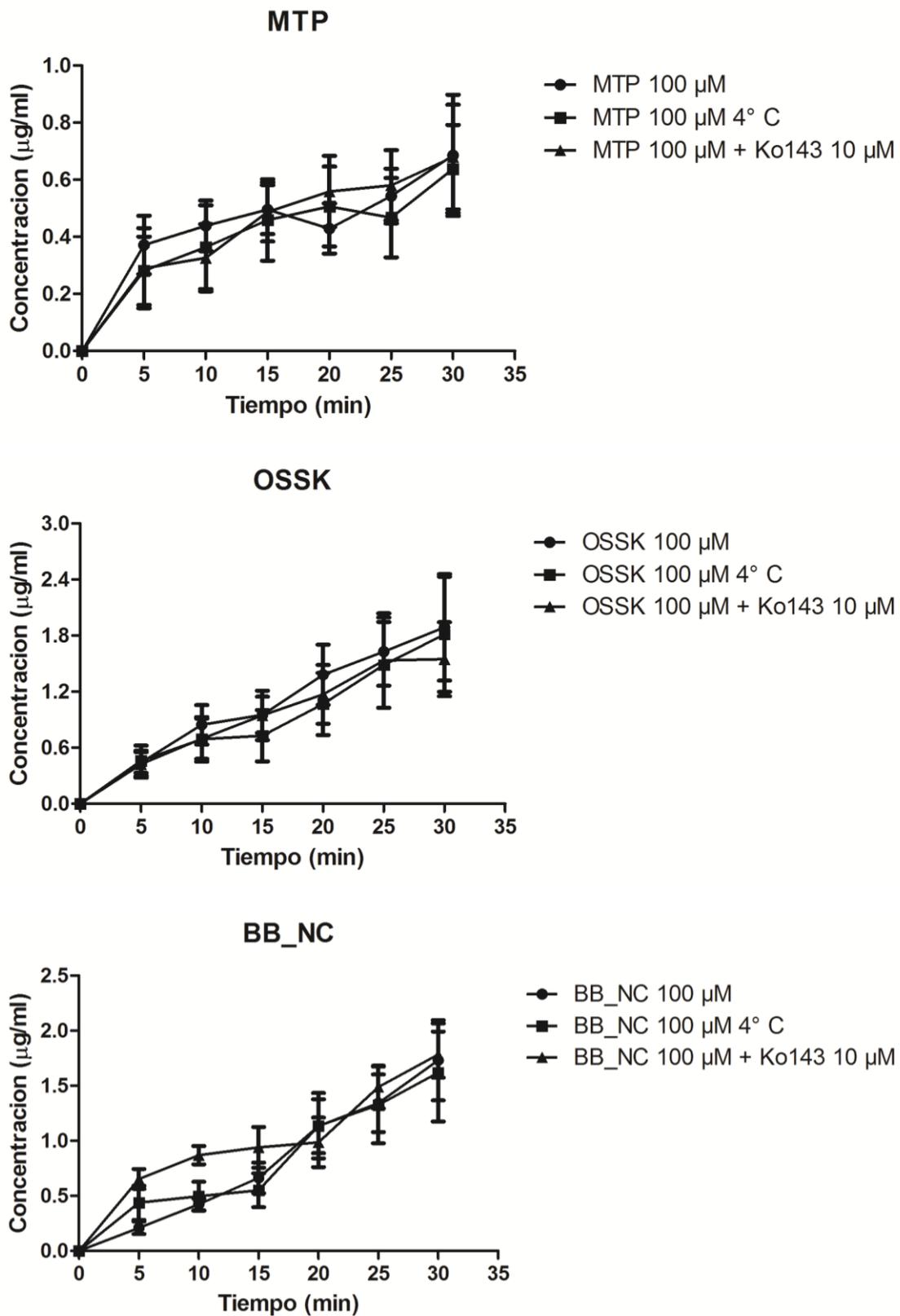
**Figura 5.9.** AUC obtenida para NF evaluada a 100 μM a 37°C en ausencia y en presencia de la droga ensayada a 10 y 100 μM (n=6). Las barras verticales indican el error estándar de la media (SEM).

Para el análisis estadístico, se aplicó el test factorial de análisis de varianza a todos los datos y no se encontraron diferencias estadísticamente significativas entre las AUC de NF en ausencia y en presencia de la droga tanto a 10 como a 100  $\mu\text{M}$  ( $p > 0.05$  en todos los casos). Los resultados indican que ninguna de las drogas evaluadas interfiere, en las concentraciones ensayadas, en el transporte de NF mediado por la BCRP de rata. Este resultado suma evidencia a la hipótesis de que las drogas no son sustrato de la BCRP, al menos no por el mismo sitio de unión o con la misma afinidad que NF en las condiciones ensayadas.

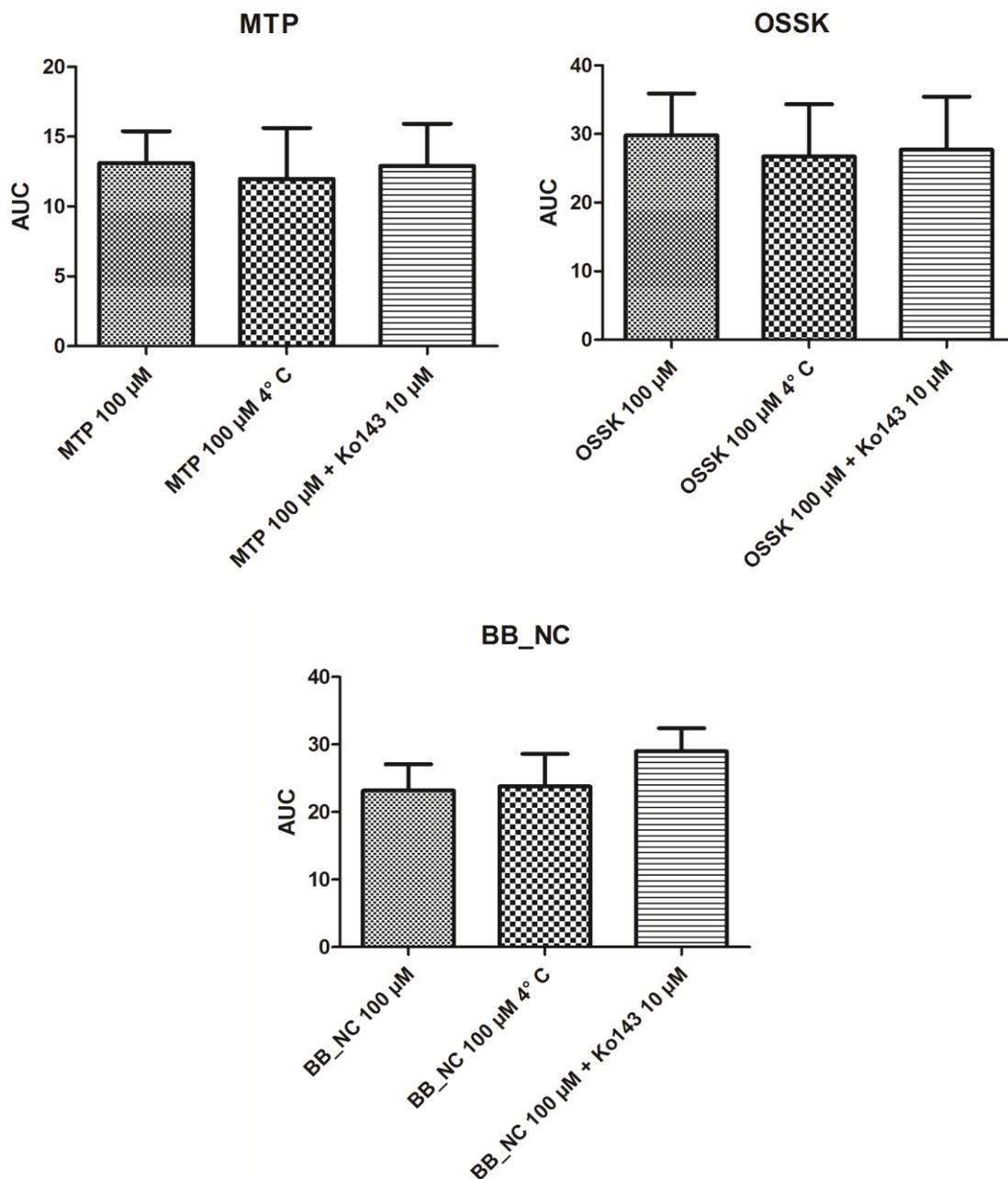
Para descartar la posibilidad de que sean un sustrato débil o que sean transportadas por un sitio de unión distinto al de NF, se ensayaron 3 de las 5 drogas (MTP, OSSK y BB\_NC) que estuvieron dentro del límite de detección del método analítico de cuantificación disponible (HPLC-UV) según el protocolo 2.

#### *5.4.5 Evaluación de las drogas según protocolo 2*

Se evaluó el transporte de cada droga tanto a 37°C como a 4°C de forma de verificar si hay transporte activo a través del tejido y a 37°C en presencia y ausencia del inhibidor Ko143 de forma de evaluar el posible transporte de las drogas por parte de la BCRP. Los resultados obtenidos se muestran en las Figuras 5.10 y 5.11.



**Figura 5.10.** Curvas obtenidas para la droga evaluada a 100 μM a 37°C y a 4°C y a 37°C en presencia de Ko143 10 μM (n=6). Las barras verticales indican el error estándar de la media (SEM).



**Figura 5.11.** AUC obtenida para las drogas evaluadas a 100 μM a 37°C y a 4°C y a 37°C en presencia de Ko143 10 μM (n=6). Las barras verticales indican el error estándar de la media (SEM).

Para el análisis estadístico, se aplicó el test factorial de análisis de varianza a todos los datos y no se encontraron diferencias estadísticamente significativas

entre las AUC de las drogas ensayadas a 100  $\mu\text{M}$  a 37°C y a 4° C y a 37 °C en presencia de Ko143 10  $\mu\text{M}$  ( $p>0.05$  en todos los casos).

De acuerdo a los resultados podemos concluir que ninguna de las drogas presenta transporte activo mediado por la BCRP de la rata en las condiciones ensayadas. Por otro lado, tampoco se observó transporte activo a través del intestino de la rata de ninguna de las tres drogas, siendo que no se encontraron diferencias entre los ensayos a 4 y 37°C.

Si bien hay que tener en cuenta la limitante de haber evaluado el modelo computacional en un modelo experimental de rata, habiendo sido desarrollado para la BCRP humana, la evidencia del protocolo 1 para las 5 drogas evaluadas, sumada a la evidencia del protocolo 2 para MTP, OSSK y BB\_NC, contribuye a confirmar la predicción hecha por el ranking promedio de los 12 mejores árboles de decisión, según la cual ninguna de las drogas sería un sustrato transportable por la BCRP.

---

## ***Referencias Capítulo V***

- [1] Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees. Belmont, CA: Chapman & Hall; 1984.
- [2] Yang H, Zhang J, Yu B, Zhao W. Statistical Methods for Immunogenicity Assessment. Hall/CRC C, editor; 2015.
- [3] Luo X, Yu JX, Li Z. 10th International Conference, ADMA 2014, Guilin, China, December 19-21, 2014, Proceedings. Advanced Data Mining and Applications: Springer International Publishing; 2014. p. XVI, 741.
- [4] Talevi A, Bellera CL, Castro EA, Bruno-Blanch LE. A successful virtual screening application: prediction of anticonvulsant activity in MES test of widely used pharmaceutical and food preservatives methylparaben and propylparaben. *J Comput Aided Mol Des.* 2007; 21(9):527-538.
- [5] Di Ianni ME, Enrique AV, Palestro PH, Gavernet L, Talevi A, Bruno-Blanch LE. Several new diverse anticonvulsant agents discovered in a virtual screening campaign aimed at novel antiepileptic drugs to treat refractory epilepsy. *J Chem Inf Model.* 2012; 52(12):3325-3330.
- [6] Couyoupetrou M, Gantner ME, Di Ianni ME, Palestro PH, Enrique AV, Gavernet L, et al. Computer-Aided Recognition of ABC Transporters Substrates and Its Application to the Development of New Drugs for Refractory Epilepsy. *Mini-Reviews in Medicinal Chemistry.* 2016; In press.
- [7] Gavernet L, Barrios IA, Cravero MS, Bruno-Blanch LE. Design, synthesis, and anticonvulsant activity of some sulfamides. *Bioorg Med Chem.* 2007; 15(16):5604-5614.



## Capítulo VI

# Conclusiones

---



Durante el presente trabajo de tesis doctoral se han desarrollado a lo largo de tres etapas de modelado una serie de modelos computacionales basados en descriptores independientes de la conformación para la identificación temprana de sustratos y no sustratos de la BCRP, proteína transportadora de eflujo vinculada a fenómenos de MDR en patologías como la epilepsia y el cáncer. Los modelos obtenidos son de fácil y rápida aplicación dado que no requieren de un análisis conformacional previo de las estructuras químicas a ser evaluadas, lo que resulta particularmente adecuado para su aplicación en campañas de cribado virtual de grandes bibliotecas químicas de una manera altamente eficiente, reduciendo el tiempo de cálculo.

Todos los modelos generados en la presente tesis se han derivado de un conjunto de datos relativamente grande y de gran diversidad estructural, el cual fue dividido en los conjuntos de entrenamiento y de prueba representativos mediante un procedimiento racional de agrupamiento, respetando el balance en el número de sustratos y no sustratos en el conjunto de entrenamiento.

De acuerdo a los resultados obtenidos en la etapa de modelado, el ERM aplicado para la obtención de la segunda serie de modelos lineales produjo mejores resultados que el método de pasos hacia adelante utilizado en la primera instancia de modelado. De la comparación de los resultados obtenidos en las tres etapas de modelado se desprende que los modelos no lineales, en este caso árboles de decisión, ajustaron mejor al problema modelado demostrando una buena capacidad predictiva en todas las instancias de validación *in silico*. Se puede destacar además la versatilidad del algoritmo J48, dado que cada descriptor puede utilizarse en más de una oportunidad (es decir, en más de un nodo), sumado a su fácil interpretación.

Por otro lado, las técnicas de fusión de datos, y en particular la combinación o ensamblado selectivo, produjeron los mejores resultados destacándose el ranking promedio y el voto promedio de los 12 mejores árboles de decisión. La amplia especificidad de sustrato de la BCRP (y, en general, de todos los transportadores ABC) hace que en este caso sea difícil encontrar un único modelo capaz de conseguir buenas tasas de predicciones tanto sobre los sustratos como los no sustratos, hecho que justificó la aplicación de estrategias más complejas, como la aplicación de métodos no lineales y la fusión de datos, constituyendo el primer

trabajo del grupo de investigación donde se aplicaron conjuntamente algoritmos genéticos y árboles de decisión.

Los resultados obtenidos durante la validación externa de los modelos y combinaciones, y particularmente la caída de la capacidad predictiva observada en la biblioteca simulada y aún más pronunciada en la biblioteca de señuelos DUD-E, demuestran la necesidad de utilizar bibliotecas de mayor tamaño y diversidad a la hora de evaluar mediante métricas de enriquecimiento (como las curvas ROC) el comportamiento *in silico* de los modelos computacionales destinados a ser aplicados en campañas de cribado virtual en grandes bibliotecas de compuestos químicos.

El ranking promedio de los 12 mejores árboles de decisión fue la combinación de modelos no lineales seleccionada para ser validada experimentalmente por ser el esquema de fusión de datos más sencillo de entre las dos combinaciones con mejor capacidad predictiva obtenidas. Dicha validación se realizó en el modelo *ex vivo* de saco intestinal evertido de rata, para lo cual se estableció en primer lugar el transporte activo de un sustrato conocido de la BCRP (NF) a través del intestino de la rata, y en segundo lugar la inhibición de ese transporte mediado por un inhibidor selectivo de la BCRP (Ko143). Cinco drogas anticonvulsivas clasificadas como no sustratos de la BCRP por la combinación de modelos computacionales fueron seleccionadas para ser evaluadas experimentalmente. Se establecieron dos protocolos: en el primero se evaluó la potencial interacción con el transportador midiendo efectos de cada droga en el transporte de NF; y en el segundo se evaluó el transporte de las drogas que no evidenciaron interacción estadísticamente significativa con la BCRP en el protocolo 1 en presencia y ausencia del inhibidor específico. Según los resultados obtenidos en el protocolo 1, ninguna de las 5 drogas a las dos concentraciones evaluadas interfirieron con el transporte activo de NF. En este punto, debe destacarse que la concentración de NF (100  $\mu\text{M}$ ) utilizada, dada la contribución de la difusión pasiva al transporte del fármaco a través del intestino evertido, podría enmascarar la interacción de los compuestos de prueba, en especial si los mismos interaccionaran débilmente con BCRP. Este efecto pesaría especialmente sobre las conclusiones respecto a la interacción o no interacción de nuestros candidatos en los dos casos (OSSL y DBS) en los que, debido a limitaciones analíticas, no se evaluó

el efecto del inhibidor específico sobre el transporte de los mismos. Idealmente, se debería repetir el protocolo 1 con menores concentraciones de NF, a fin de ganar sensibilidad en la detección de posibles interacciones entre nuestros candidatos y la BCRP.

En cuanto al protocolo 2, sólo se pudieron evaluar 3 drogas (MTP, OSSK y BB\_NC) que, para los niveles ensayados, lograron concentraciones por encima del límite de detección del método analítico de cuantificación disponible (HPLC-UV). En esta etapa, ninguna de las tres drogas ensayadas evidenció transporte activo por parte de la BCRP a través del intestino de la rata en las condiciones experimentales utilizadas. La evidencia del protocolo 1 para las 5 drogas evaluadas, sumada a la evidencia del protocolo 2 para MTP, OSSK y BB\_NC, parece sustentar las predicciones hechas por el ranking promedio de los 12 mejores árboles de decisión (con las limitaciones mencionadas), según las cuales ninguna de las drogas sería un sustrato transportable por la BCRP. Debe destacarse también como una limitante de la validación experimental el hecho de haber utilizado un modelo de rata, siendo que el modelo computacional está destinado al reconocimiento de sustratos y no sustratos de la BCRP humana y por ese motivo fue desarrollado a partir de un conjunto de datos especialmente filtrado para tal fin, considerando que puede haber cambios en la especificidad de sustrato dependientes de la especie.

Estos resultados contribuyen a la validación experimental de la capacidad predictiva de la combinación de modelos computacionales desarrollada, sugiriendo que tales modelos constituyen una herramienta potencialmente valiosa para ser utilizada como filtro ADME *in silico* y ayudar en campañas de diseño de fármacos asistido por computadora, e ilustrando el potencial de la estrategia propuesta para el desarrollo de nuevos medicamentos que permitan superar problemas de MDR mediada por la BCRP. Como una ventaja adicional, los modelos permiten la predicción de potenciales interacciones medicamentosas vinculadas a la administración simultánea de dos o más fármacos sustratos, inhibidores o inductores de dicho transportador.

---

**Detalle de las publicaciones realizadas, becas obtenidas y presentaciones a congresos durante el período en el que se desarrolló la presente Tesis doctoral:**

Durante este trabajo de tesis se han realizado 7 presentaciones a congresos (2 nacionales y 5 internacionales), 3 publicaciones originales en revistas científicas con referato, 1 trabajo en redacción, además de 1 capítulo de libro y 5 revisiones.

En el año 2012 se ha obtenido una beca de la TWAS-ROLAC (The Academy of Sciences for the Developing World-Regional Office for Latin America and the Caribbean) y The Center for Special Studies and Programs (CSSP) of the Bibliotheca Alexandrina (BA) para la asistencia a la Academy of Sciences for the Developing World/BioVisionAlexandria.NXT 2012 Program “Scientific Innovation in the Developing World: from Theory to Practice” y a BioVisionAlexandria 2012 International Biennial Conference “New Life Sciences: Linking Science to Society” en la Biblioteca de Alejandría, Egipto. En 2013 la Sociedad Iberoamericana de Bioinformática otorgó una beca para asistir al 4to. Congreso Argentino de Bioinformática y Biología Computacional (4CAB2C) y a la 4ta. Conferencia Internacional de la Sociedad Iberoamericana de Bioinformática (SolBio). En el año 2014 la International League Against Epilepsy (ILAE) y The International Bureau for Epilepsy (IBE) otorgaron una beca para la asistencia al VIII Congreso Latinoamericano de Epilepsia; en el mismo año, The Institut Pasteur de Montevideo otorgó una beca para la asistencia al Workshop Internacional “Modern Approaches in Drug Discovery for neglected infectious diseases”.

A continuación se detallan las publicaciones realizadas en el marco de la tesis doctoral:

**Artículos:**

- ⊗ Melisa Edith Gantner, Mauricio Emiliano Di Ianni, María Esperanza Ruiz, Alan Talevi y Luis E. Bruno-Blanch. Development of Conformation Independent Computational Models for the Early Recognition of Breast Cancer Resistance Protein substrates. BioMed Research International (hasta 2013, Journal of Biomedicine and Biotechnology). Volumen 2013, Article ID 863592, 12 pages, 2013. Con referato.

- ⊗ Mauricio E. Di Ianni, Melisa E. Gantner, María E. Ruiz, Eduardo A. Castro, Luis E. Bruno-Blanch y Alan Talevi. Systematic comparison of the performance of different 2D and 3D ligand-based virtual screening methodologies to discover anticonvulsant drugs. *Combinatorial Chemistry & High Throughput Screening*. Octubre de 2014. Con referato.
- ⊗ Melisa E. Gantner, Lucas N. Alberca, Andrew Mercader, Luis E. Bruno-Blanch y Alan Talevi. Integrated Application of Enhanced Replacement Method and Ensemble Learning for the Prediction of BCRP/ABCG2 Substrates. *Current Bioinformatics*. 2014. Con referato. En prensa.
- ⊗ Melisa E. Gantner, Roxana N. Peroni, Juan F. Morales, María L. Villalba, María E. Ruiz y Alan Talevi. " Development and validation of a computational model ensemble for the early detection of BCRP/ABCG2 substrates during the drug design stage". Enviado para su publicación.

#### **Revisiones:**

- ⊗ Alan Talevi, Carolina L. Bellera, Mauricio E. Di Ianni, Melisa E. Gantner, Luis E. Bruno-Blanch y Eduardo A. Castro. CNS Drug development – Lost in translation?. *Mini-Reviews in Medicinal Chemistry*. 12 (10). 2012. pp.959-970. Con referato.
- ⊗ Alan Talevi, Melisa E. Gantner y María E. Ruiz. Applications of Nanosystems to Anticancer Drug Therapy (Part I. Nanogels, Nanospheres, Nanocapsules). *Recent Patents on Anti-Cancer Drug Discovery*. 9 (1). 2014. pp. 83-98. Con referato.
- ⊗ María E. Ruiz, Melisa E. Gantner y Alan Talevi. Applications of Nanosystems to Anticancer Drug Therapy (Part II. Dendrimers, Micelles, Lipid-based Nanosystems). *Recent Patents on Anti-Cancer Drug Discovery*. 9 (1). 2014. pp. 99-128. Con referato.
- ⊗ Carolina L. Bellera, Melisa E. Gantner, María E. Ruiz y Alan Talevi. Recent advances on nanotechnology applications to cancer drug therapy. *Journal of Cancer Research Updates*. 2 (3). 2013. pp. 151-185. Con referato.
- ⊗ Manuel Couyoupetrou, Melisa E. Gantner, Mauricio E. Di Ianni, Pablo H. Palestro, Andrea V. Enrique, Luciana Gavernet, María E. Ruiz, Guido Pesce, Luis E. Bruno-Blanch y Alan Talevi. Computer-Aided Recognition of ABC

Transporters Substrates and Its Application to the Development of New Drugs for Refractory Epilepsy. Mini-Reviews in Medicinal Chemistry. 2015. Con referato. En prensa.

**Capítulos de Libro:**

- ∅ María E. Ruiz, Melisa E. Gantner y Alan Talevi. Recent Patents on Polymeric Nanosystems Applications for Anticancer Drug Delivery. En: Topics in Anti-Cancer Research. Volumen 4, 2015, pp. 524-601. Editores: Atta-ur-Rahman y Khurshid Zaman. Bentham Science Publishers.

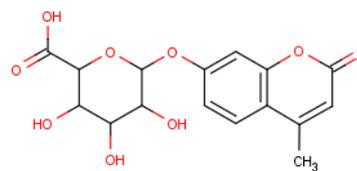
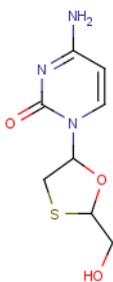
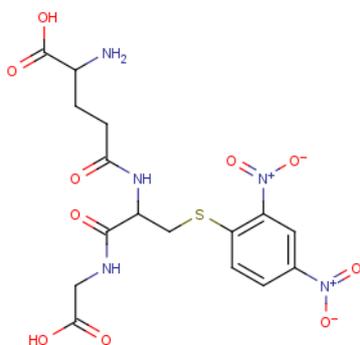
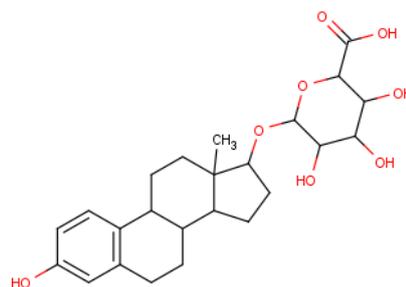
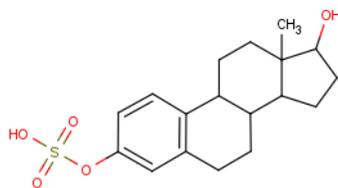
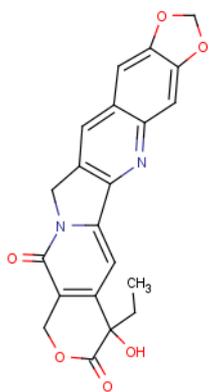
# Anexo I

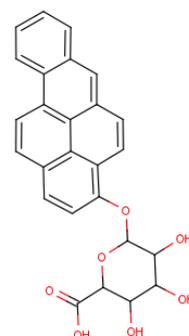
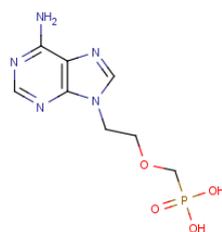
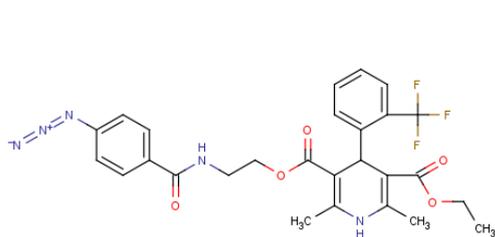
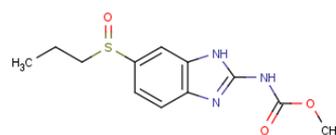
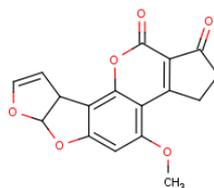
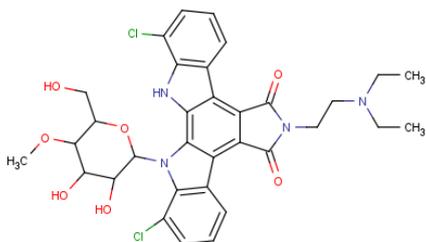
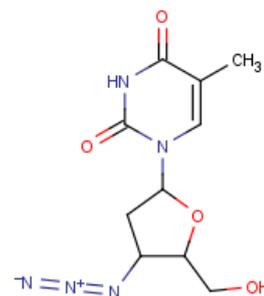
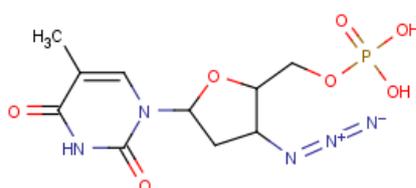
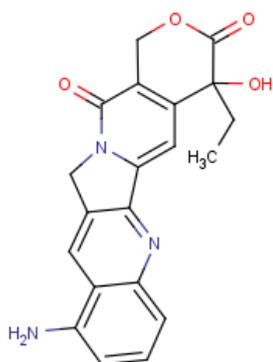
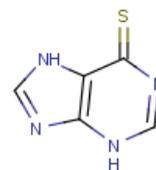
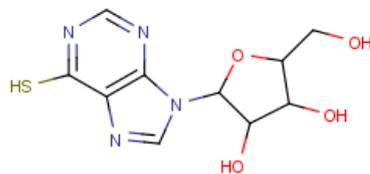
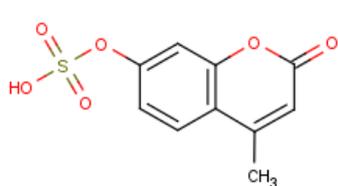
---

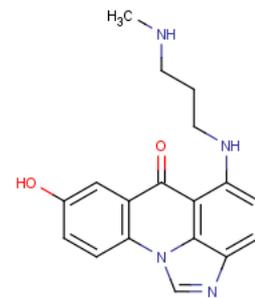
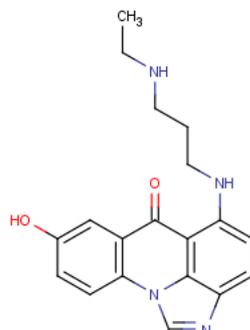
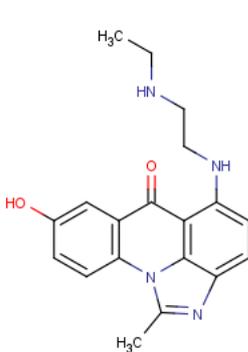
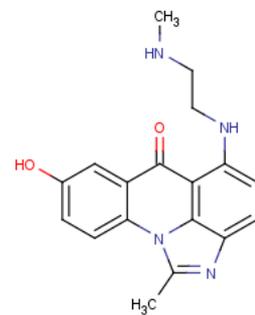
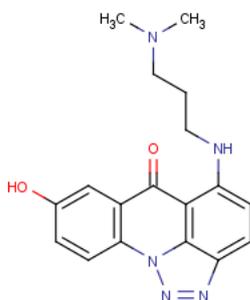
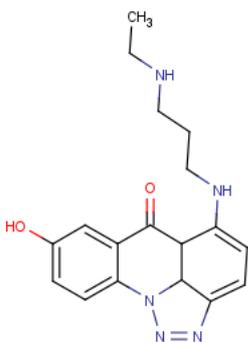
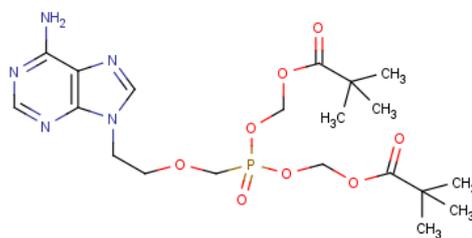
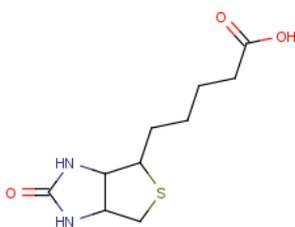
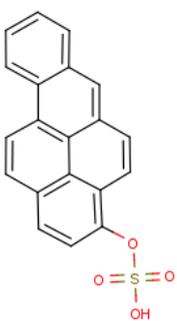
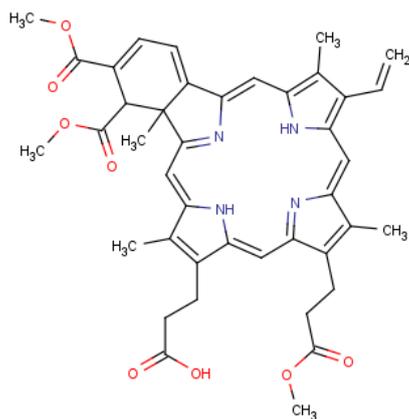
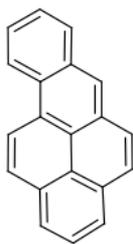
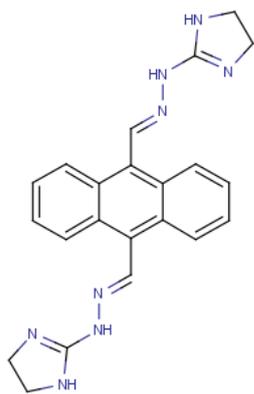


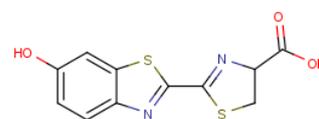
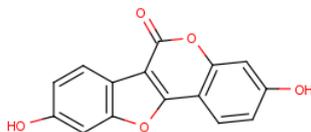
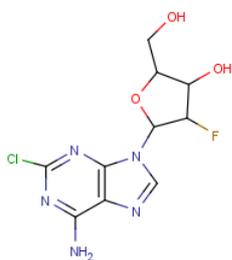
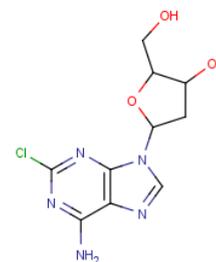
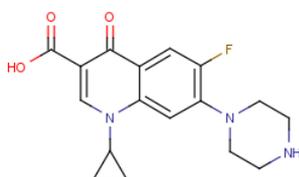
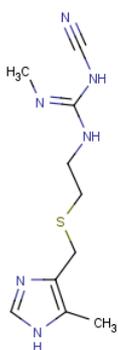
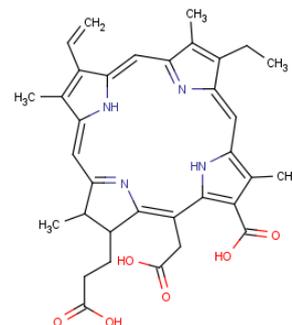
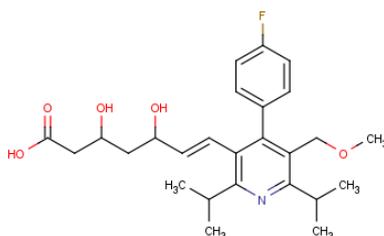
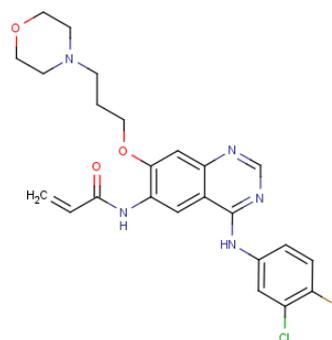
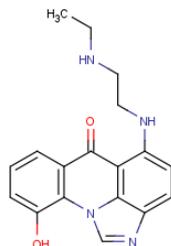
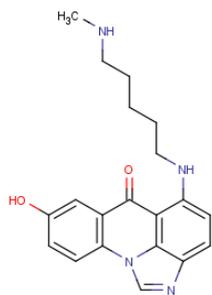
## A.1.1 Estructuras químicas de los compuestos del conjunto de datos

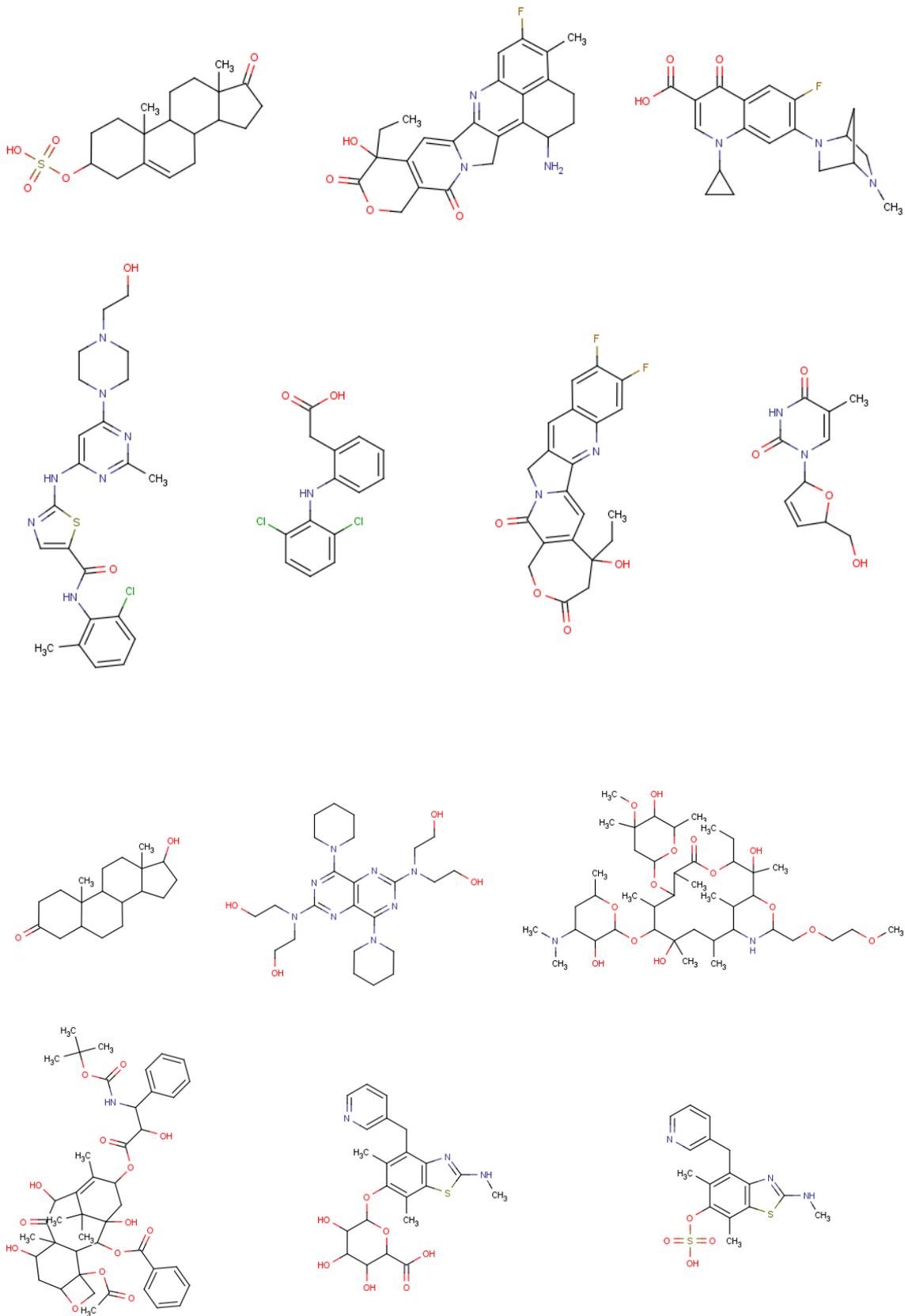
### A.1.1.1 Estructuras químicas de los sustratos de la BCRP del conjunto de entrenamiento



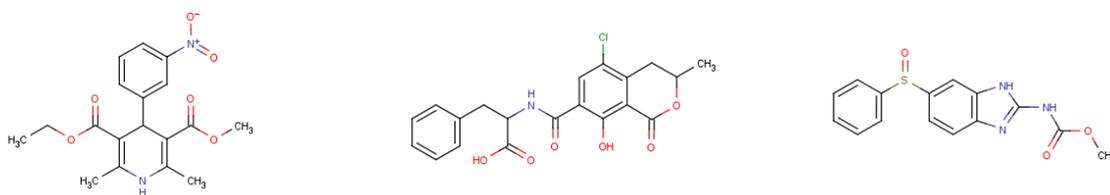
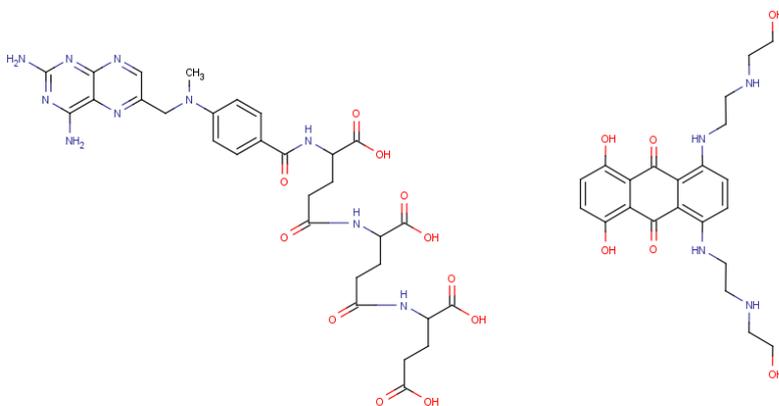
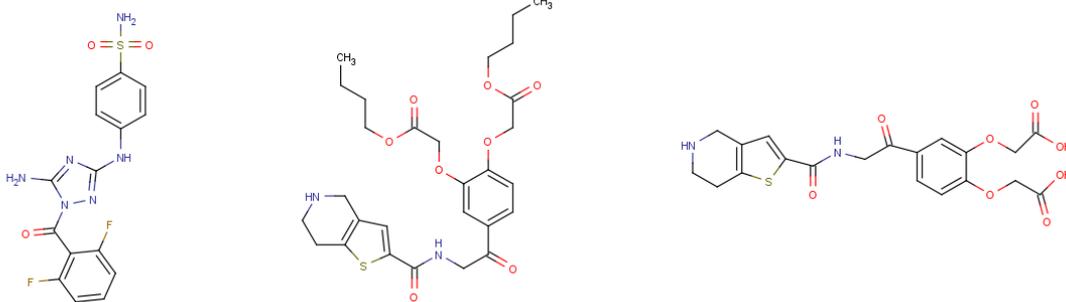
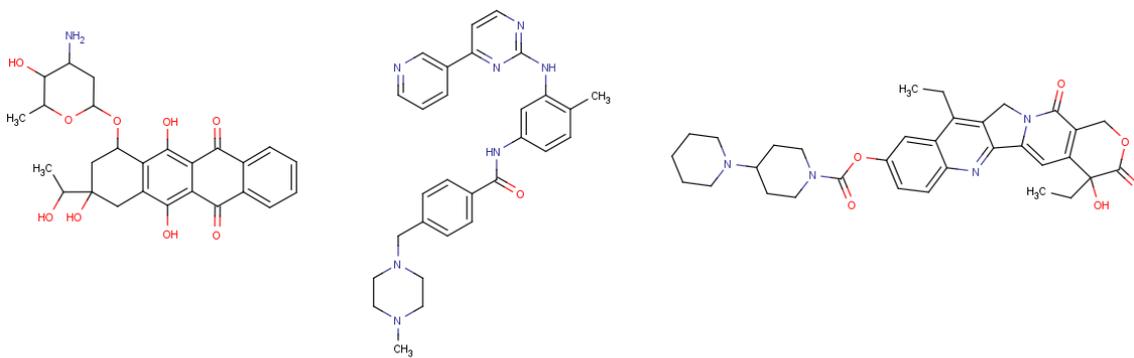


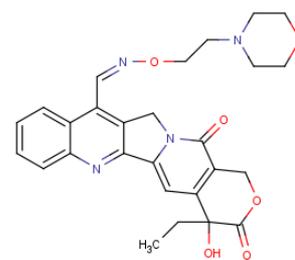
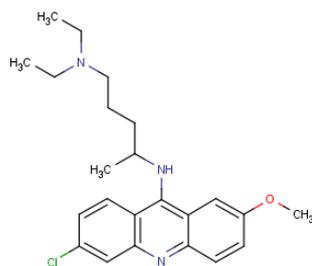
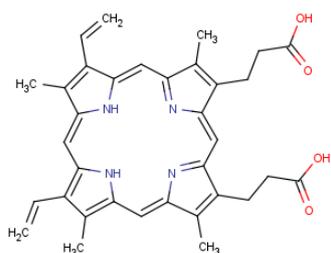
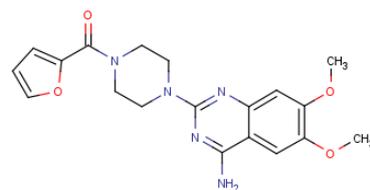
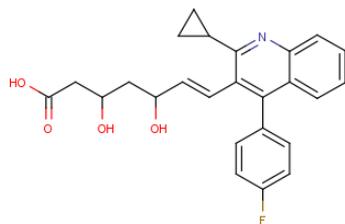
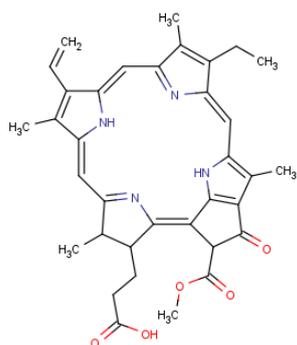




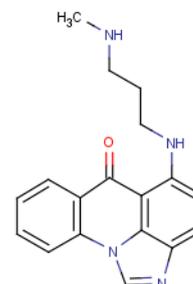
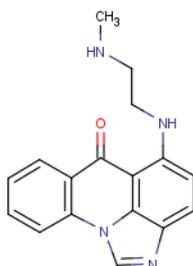
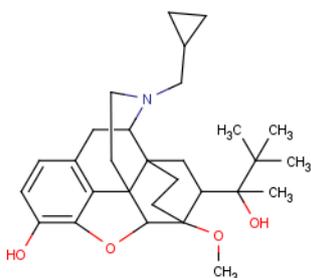
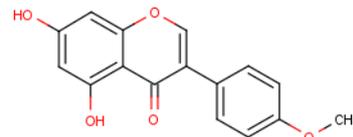
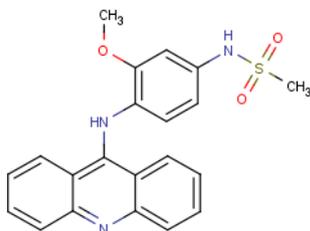
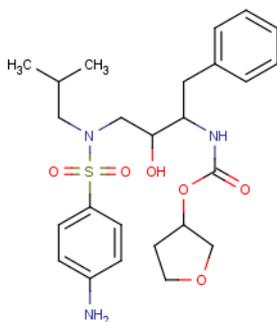
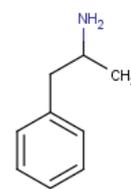
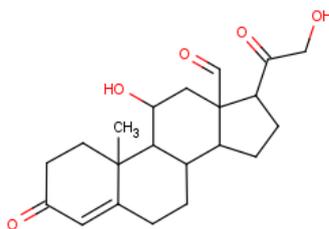
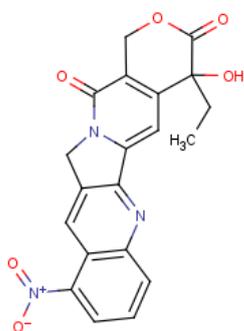
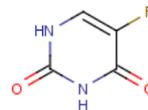
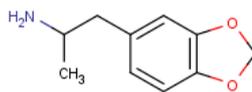
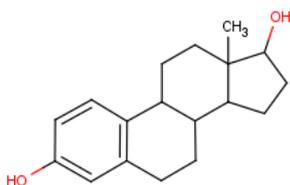




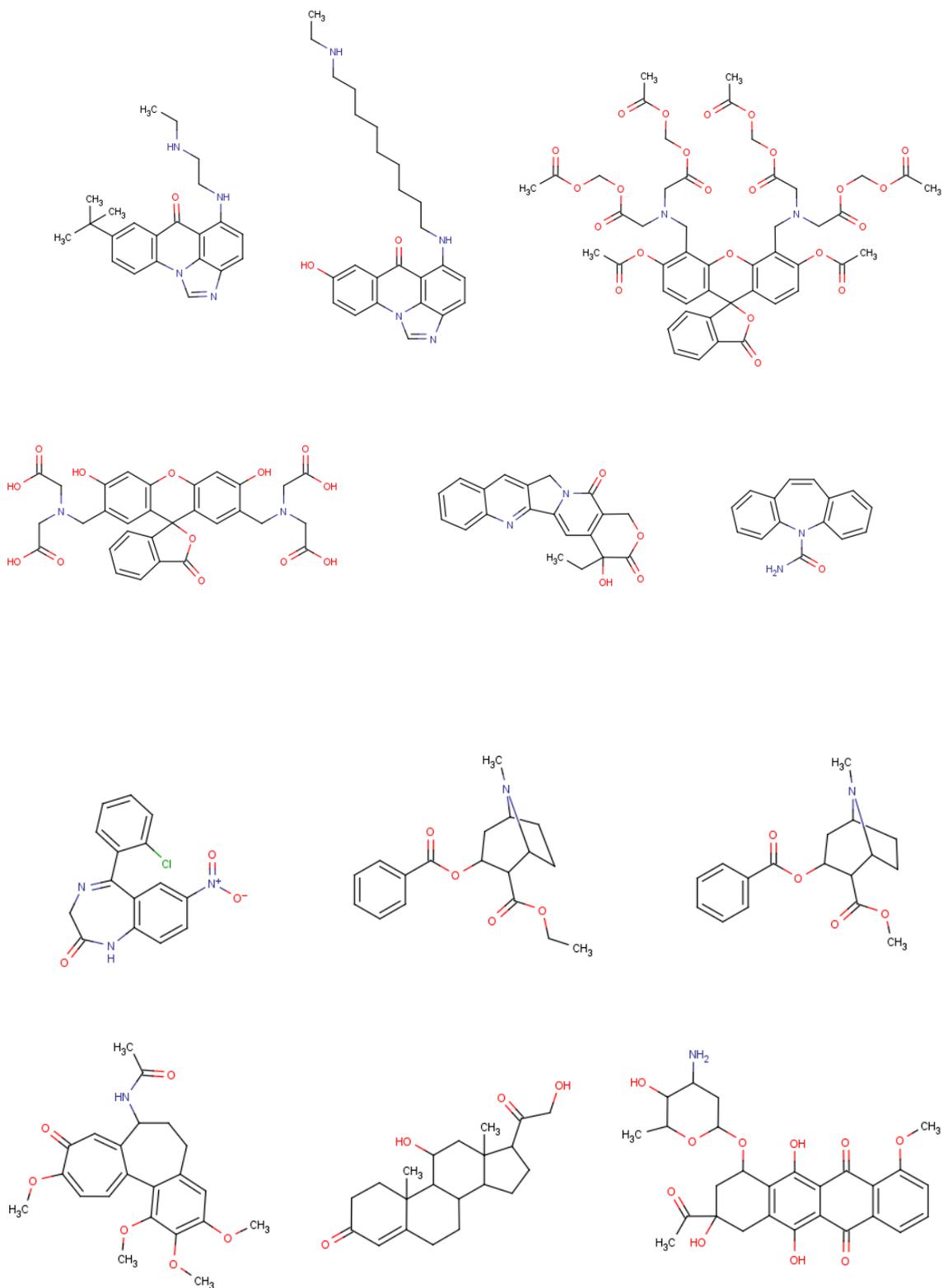


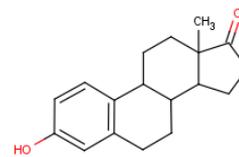
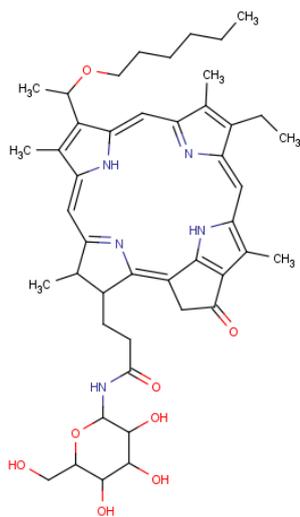
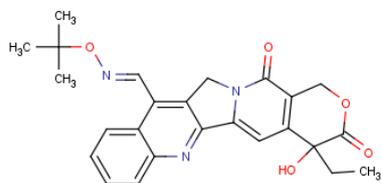
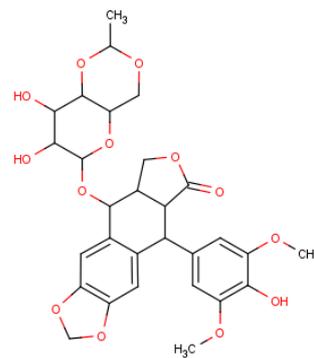
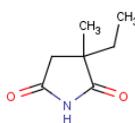
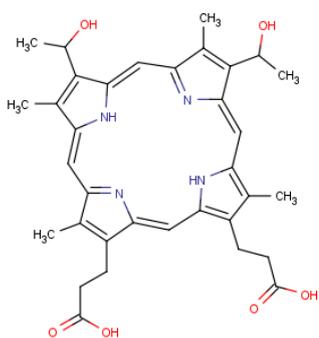
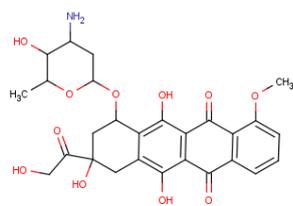
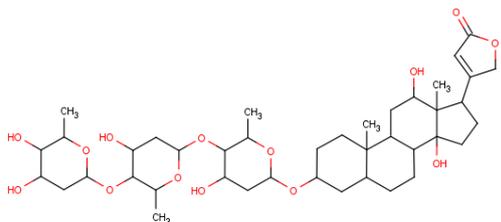
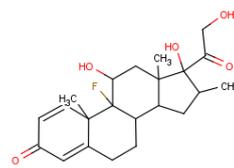
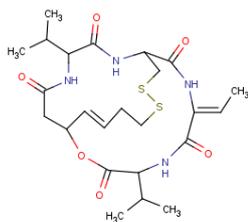
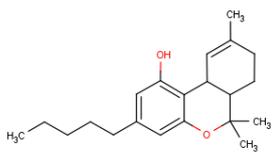


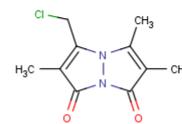
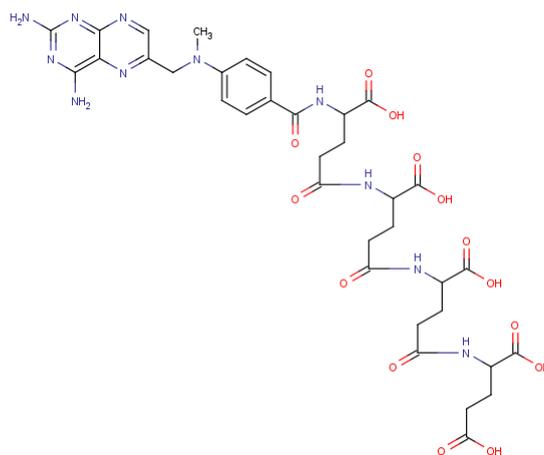
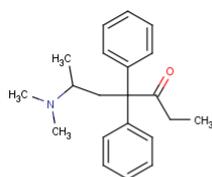
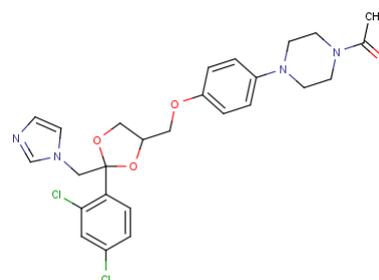
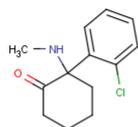
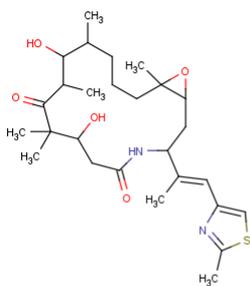
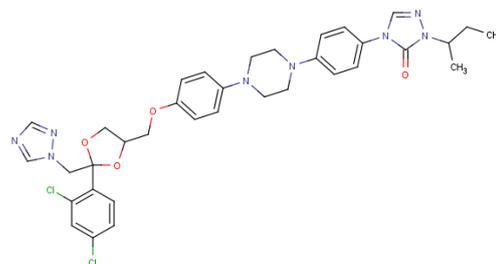
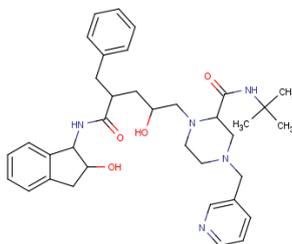
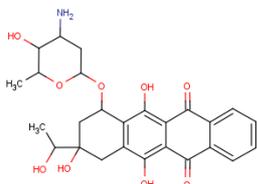
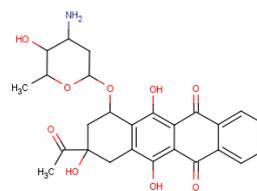
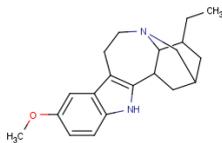
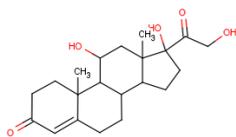
*A.1.1.2 Estructuras químicas de los no sustratos de la BCRP del conjunto de entrenamiento*

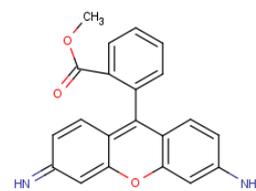
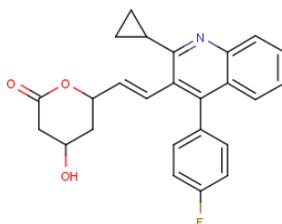
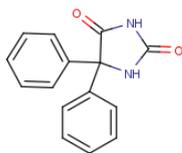
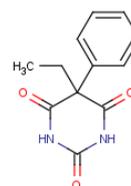
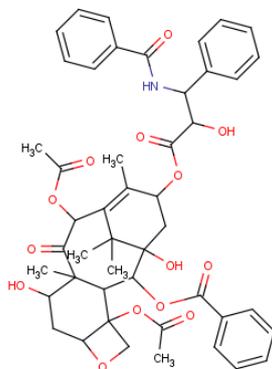
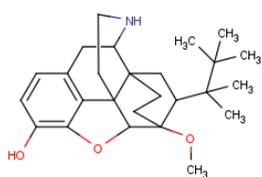
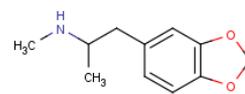
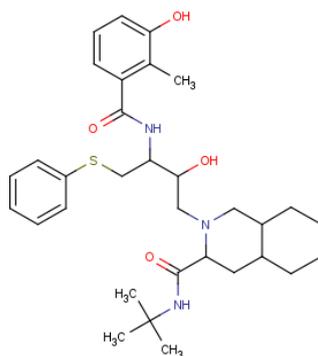
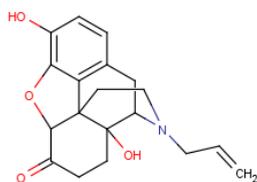
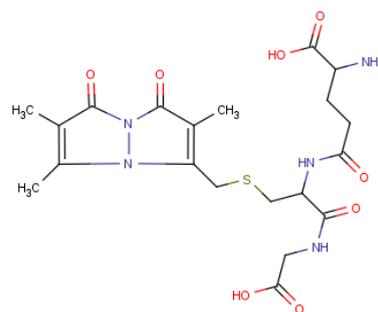
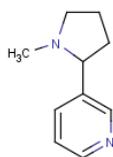
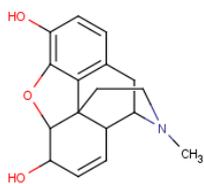






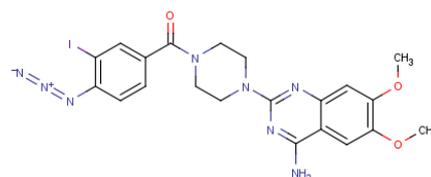
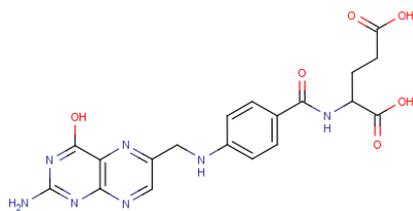
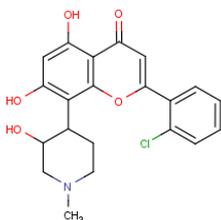
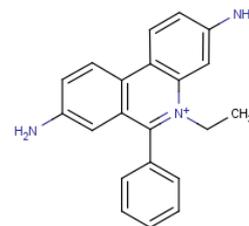
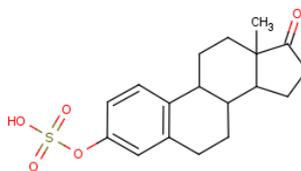
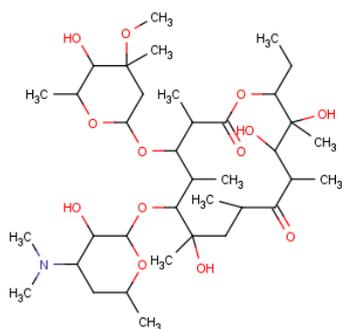
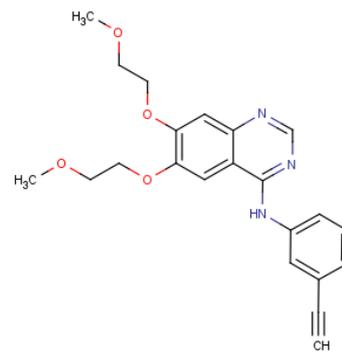
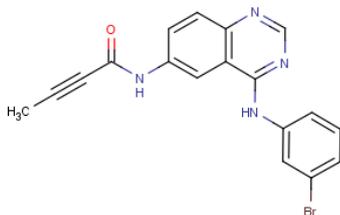
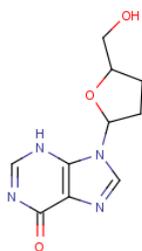


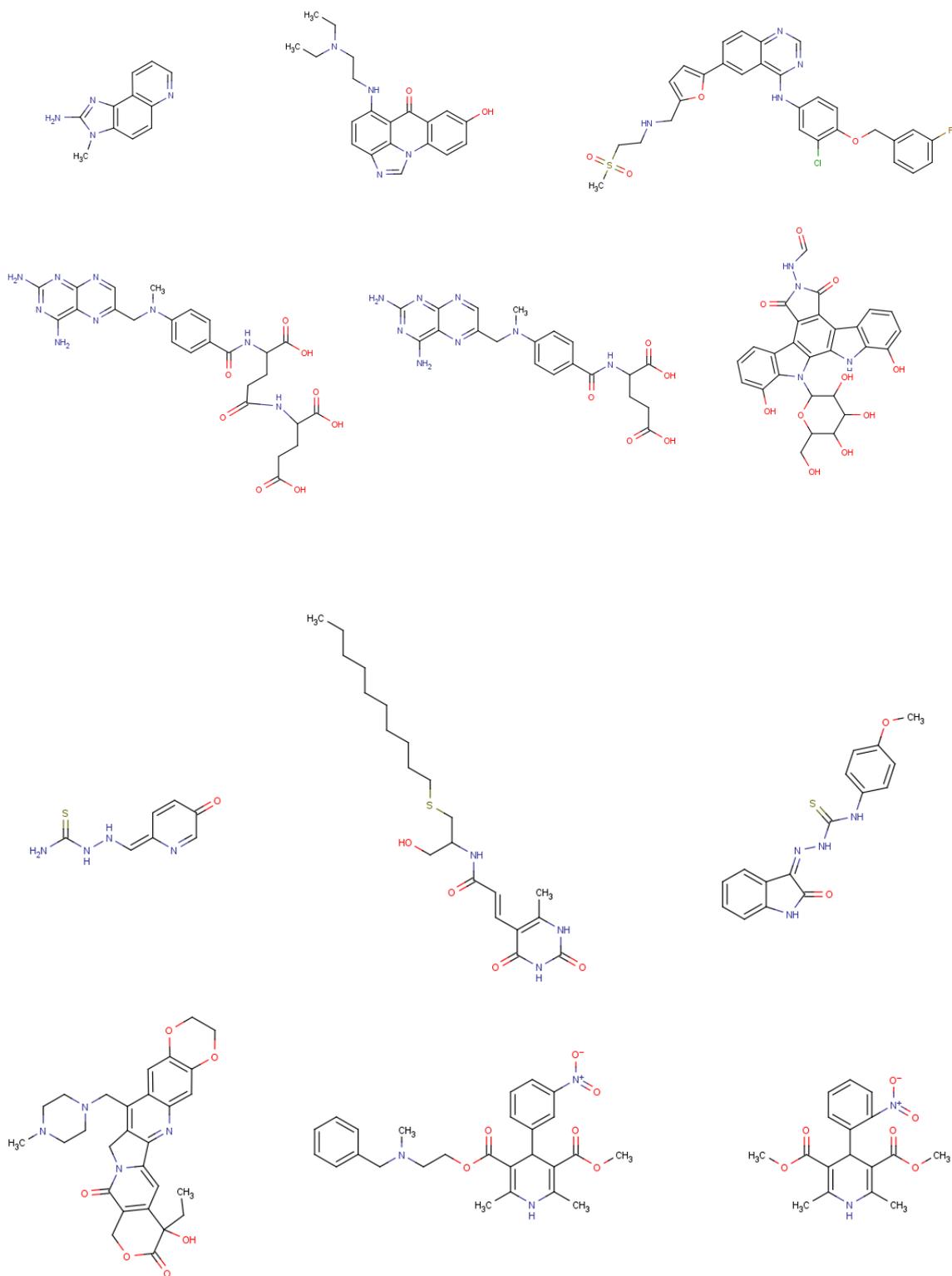


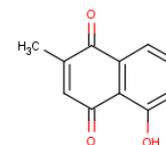
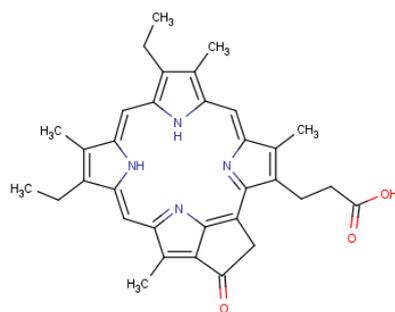
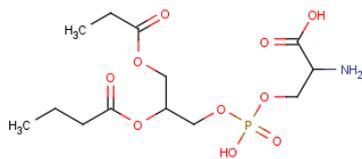
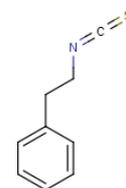
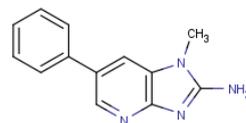
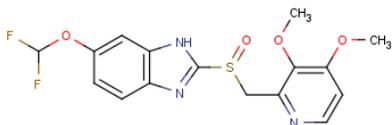
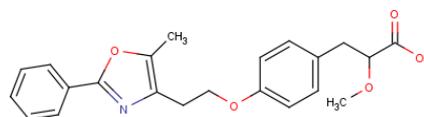
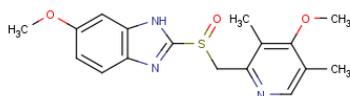
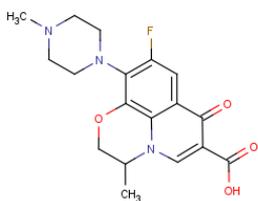
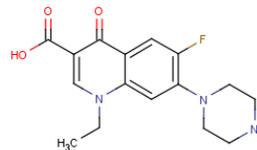
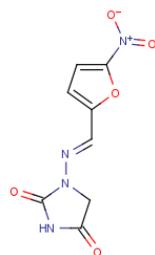
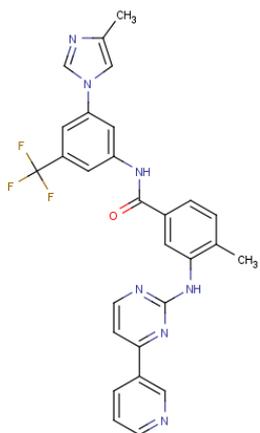


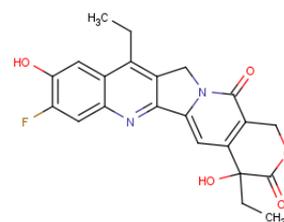
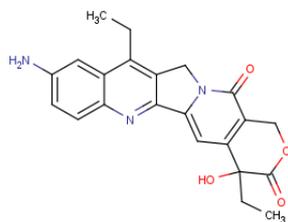
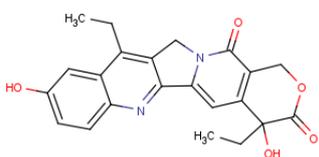
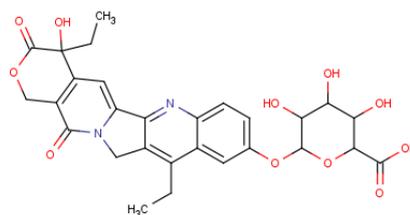
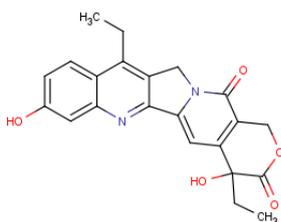
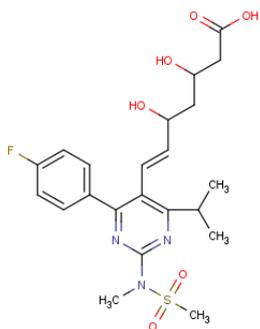
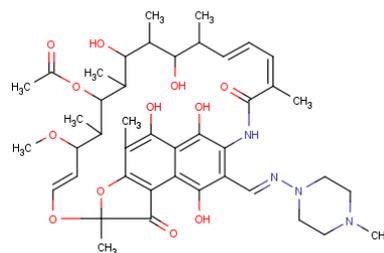
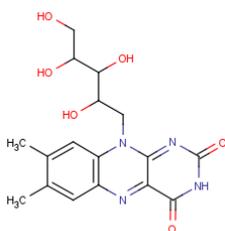
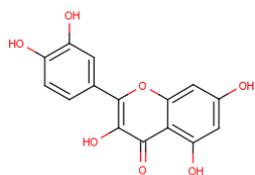
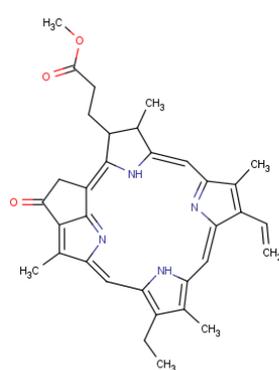
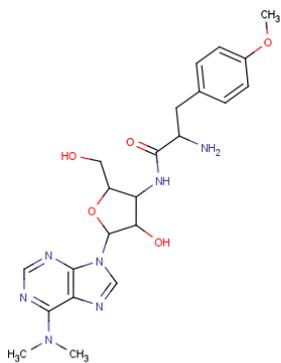
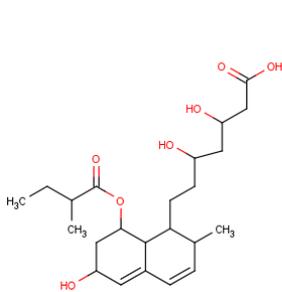


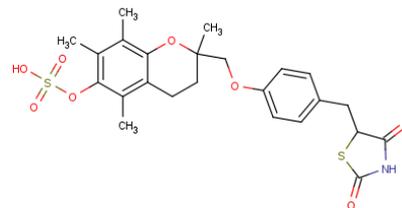
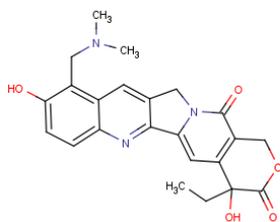
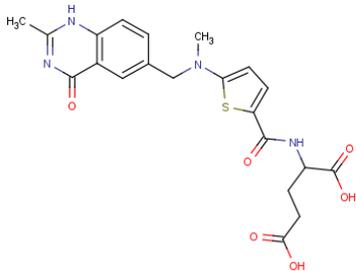
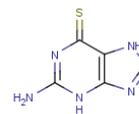
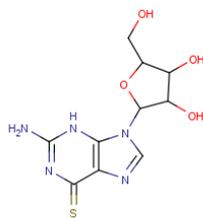
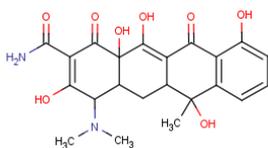
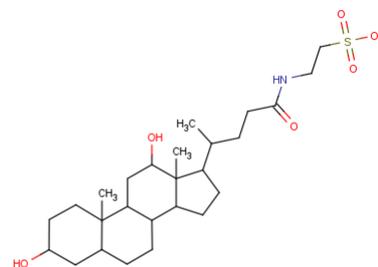
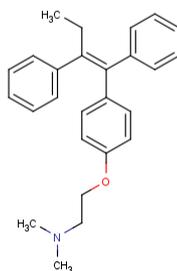
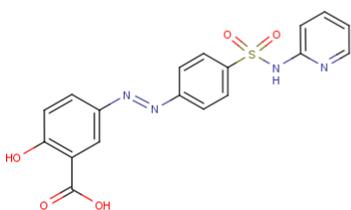
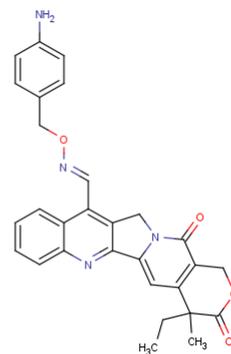
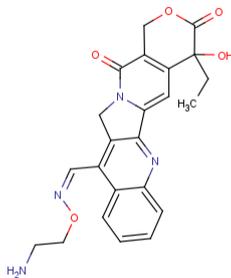
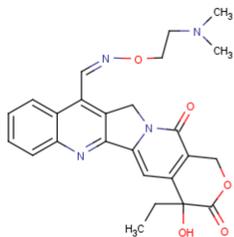


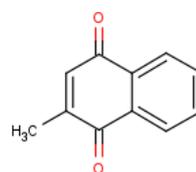
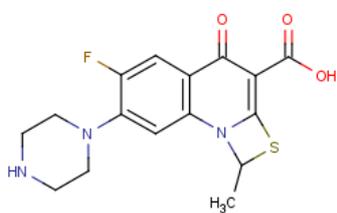
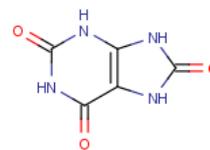
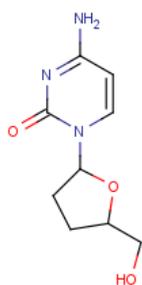
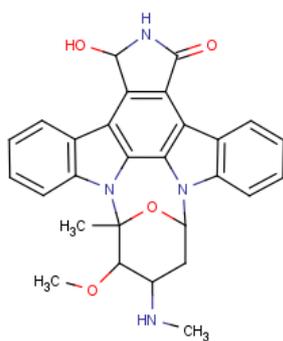




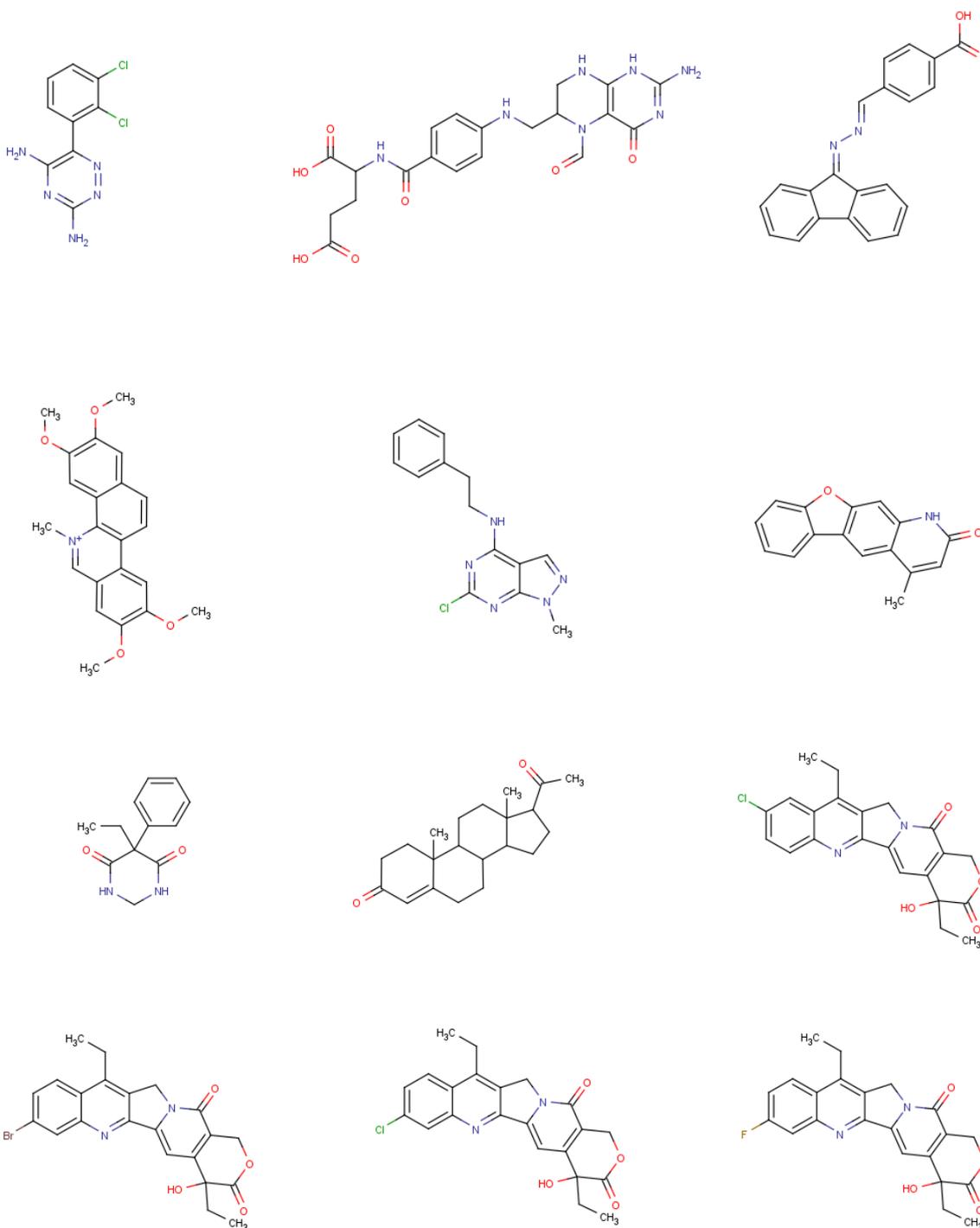


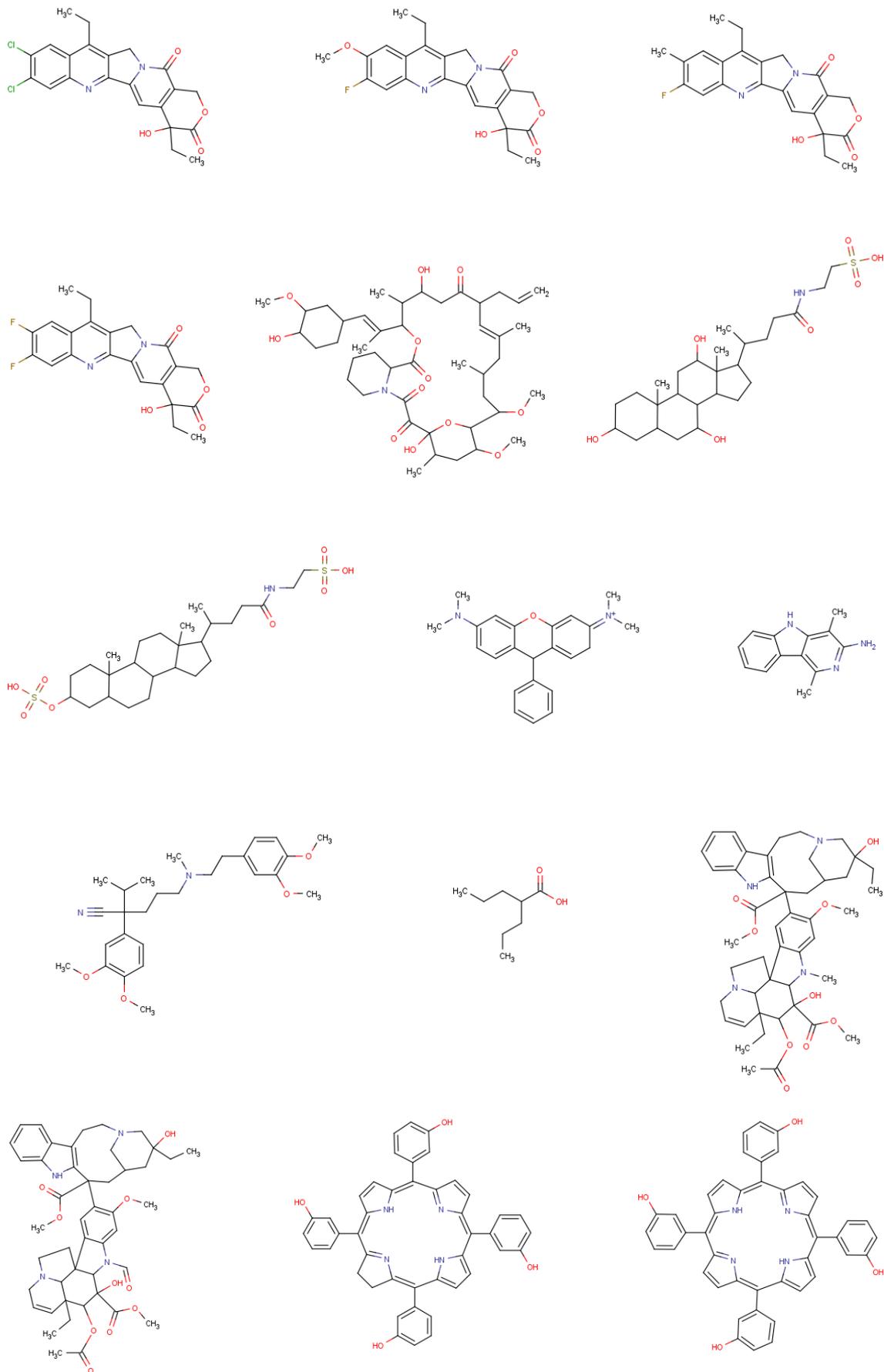




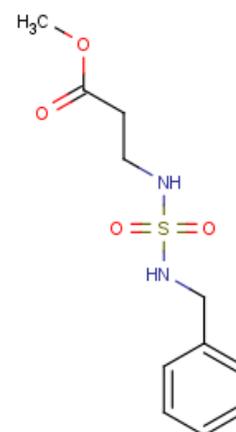
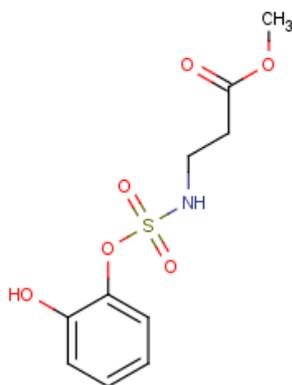
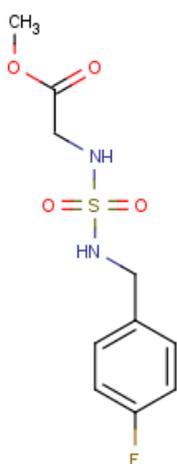
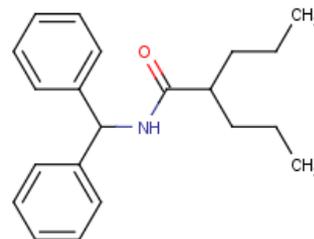
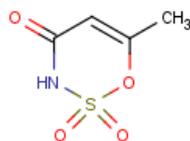
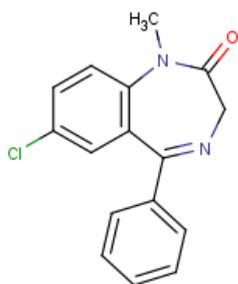
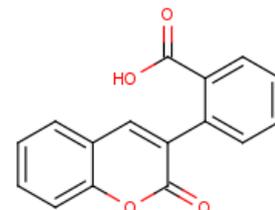
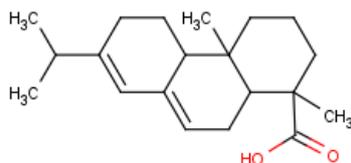
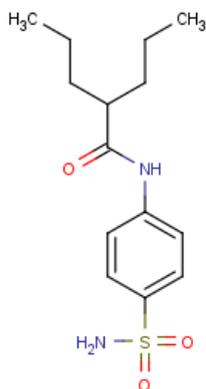


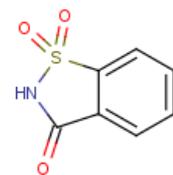
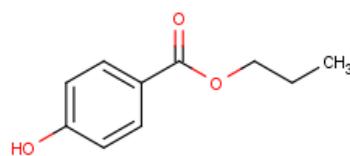
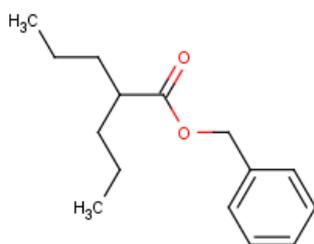
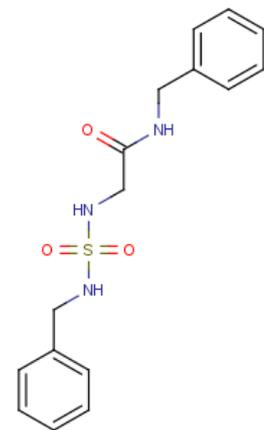
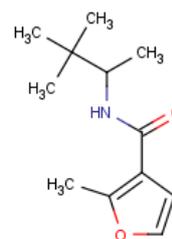
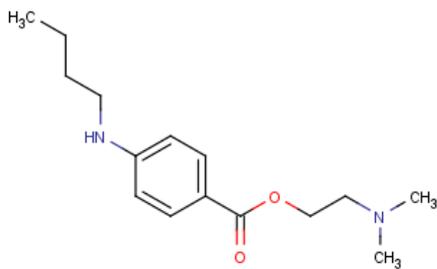
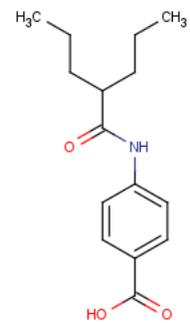
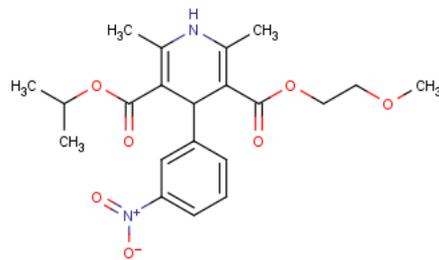
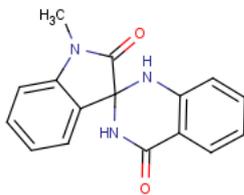
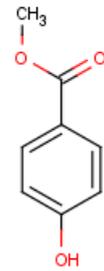
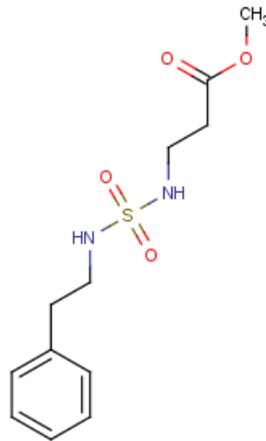
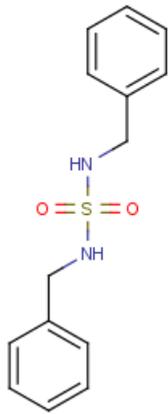
*A.1.1.4 Estructuras químicas de los no sustratos de la BCRP del conjunto de prueba*





**A.1.2 Estructuras químicas de la base de datos de 21 compuestos anticonvulsivos identificados por el LIDeB sobre la que se aplicó la mejor combinación de modelos obtenida**





# Anexo II

---



### A.2.1 Ecuaciones de la segunda serie de modelos lineales

Se presentan las ecuaciones de los 10 mejores modelos lineales obtenidos por ERM en la segunda instancia de modelado. Se encuentran ordenados de la mayor a la menor AUC ROC para el conjunto de prueba. Se mantuvo para los descriptores moleculares la nomenclatura del Dragon 6.0.

M438:

$$Df = -3.9626 - 0.9292nR07 + 0.01945D/Dtr04 + 0.55150piPC09 \\ + 1.70276ATSC2e + 1.64049GATS1e + 2.65398SpMax2_Bh(i) \\ + 0.32361SpMax2_Bh(s) - 1.18587SM04_EA(ed) \\ - 1.55021 Eig10_EA(dm) - 0.70658Uc + 1.46549 DLS_01$$

M380:

$$Df = 7.7329 + 3.6061 IDE - 0.9949 WiA_Dz(m) - 13.9889 SpPosA_B(v) \\ + 0.1937 SpMax2_Bh(s) + 0.1854 SaaaC + 0.8424 SdS \\ + 0.1223 CATS2D_05_DA + 0.5065 CATS2D_01_NL \\ - 1.1134 B10[C - C] + 1.1267 DLS_01$$

M107:

$$Df = -7.4748 + 3.3877 EE_B(e) - 0.01476 P_VSA_i_2 - 1.9473 SpDiam_EA \\ - 0.6035 nRCONHR - 0.7049 B10[C - C] + 1.6271 DLS_01$$

M471:

$$Df = -13.1926 - 0.1154 SRW09 + 3.7279 IDE - 1.0708 WiA_Dz(e) \\ + 0.4491 SpMax2_Bh(s) + 9.7761 Eta_betaS_A + 0.2753 C - 034 \\ - 0.4853 CATS2D_08_DP + 0.2226 CATS2D_07_NL - 1.5344 B02[C \\ - O] + 0.3689 B08[N - O] - 0.5810 B10[C - C] - 0.7177 Ui$$

M356:

$$Df = 5.8284 + 0.0660 N\% + 1.3060 nR04 + 0.4651 SpMax2_Bh(s) \\ - 0.6772 SM11_EA + 1.5036 SM13_AEA(bo) - 0.0843 nCs \\ - 0.6392 nRCONHR + 0.1858 NaaaC + 0.2781 CATS2D_07_NL \\ - 0.965219 Ui$$

M61:

$$Df = 0.9237 + 0.2780 \text{ SpMax2\_Bh}(s) - 3.2111 \text{ SpMaxA\_EA}(ed) \\ - 0.6101 \text{ nRCONHR} + 0.2969 \text{ CATS2D\_07\_NL} - 0.0662 \text{ F01}[C - C]$$

M203:

$$Df = 12.2138 + 4.3538 \text{ IDE} - 1.2489 \text{ SpMaxA\_Dz}(e) - 16.4320 \text{ SpPosA\_B}(v) \\ - 0.2572 \text{ nArOR} + 4.0514 \text{ S} - 108 + 0.1243 \text{ NaaaC} \\ + 0.2117 \text{ CATS2D\_07\_NL} - 0.9246 \text{ B10}[C - C]$$

M289:

$$Df = 5.1028 + 0.0643 \text{ N\%} + 0.4580 \text{ SpMax2\_Bh}(s) - 0.6278 \text{ SM11\_EA} \\ + 1.4333 \text{ Eig05\_EA} - 0.0789 \text{ nCs} - 0.6267 \text{ nRCONHR} \\ + 0.1646 \text{ NaaaC} + 0.2766 \text{ CATS2D\_07\_NL} - 0.8766 \text{ Ui}$$

M36:

$$Df = -5.8113 + 2.0526 \text{ SM5\_B}(e) - 0.0136 \text{ P\_VSA\_i\_2} - 1.6891 \text{ SpDiam\_EA} \\ - 0.6538 \text{ nRCONHR}$$

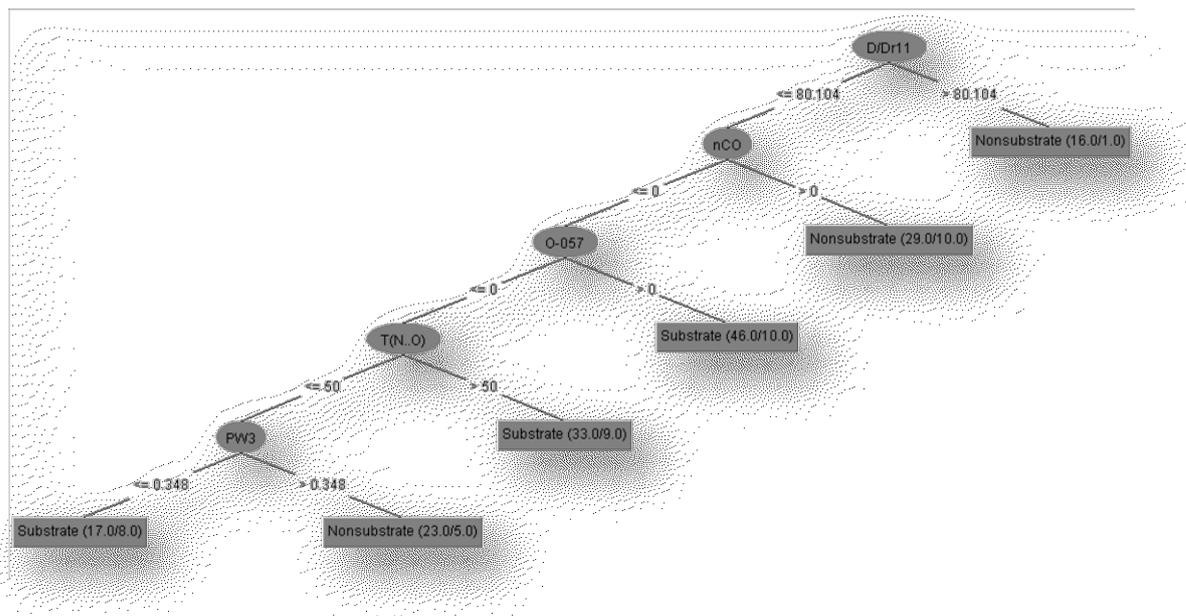
M446:

$$Df = 10.9699 + 1.9349 \text{ nR04} - 5.8532 \text{ SpMax\_A} + 0.8235 \text{ ATSC2e} \\ + 0.2316 \text{ SpMax2\_Bh}(s) - 0.0123 \text{ P\_VSA\_i\_2} + 1.2408 \text{ Eig05\_EA} \\ - 0.6502 \text{ nRCONHR} + 0.2423 \text{ NaaaC} - 0.4433 \text{ B03}[O - O] \\ - 0.3177 \text{ F09}[C - S] + 0.7895 \text{ cRo5}$$

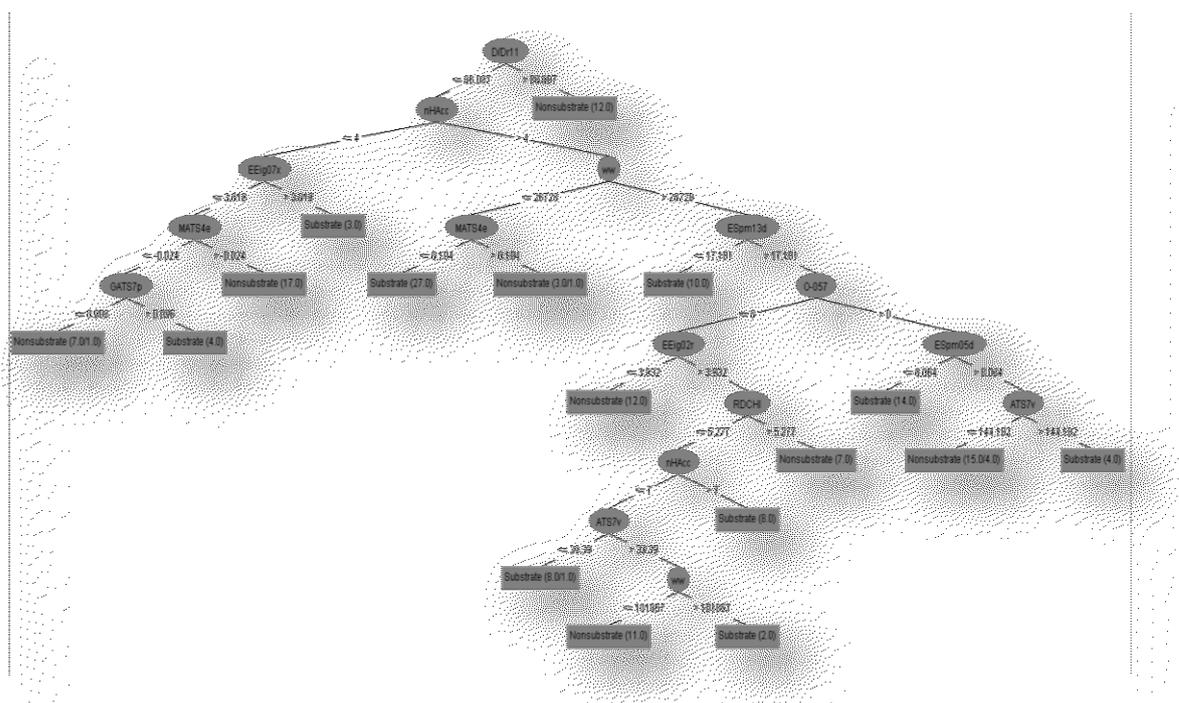
## A.2.2 Diagramas de los 12 árboles de decisión obtenidos

Se presentan los diagramas de los 12 mejores árboles de decisión obtenidos en la tercera instancia de modelado. Se mantuvo para los descriptores moleculares la nomenclatura del Dragon 4.0.

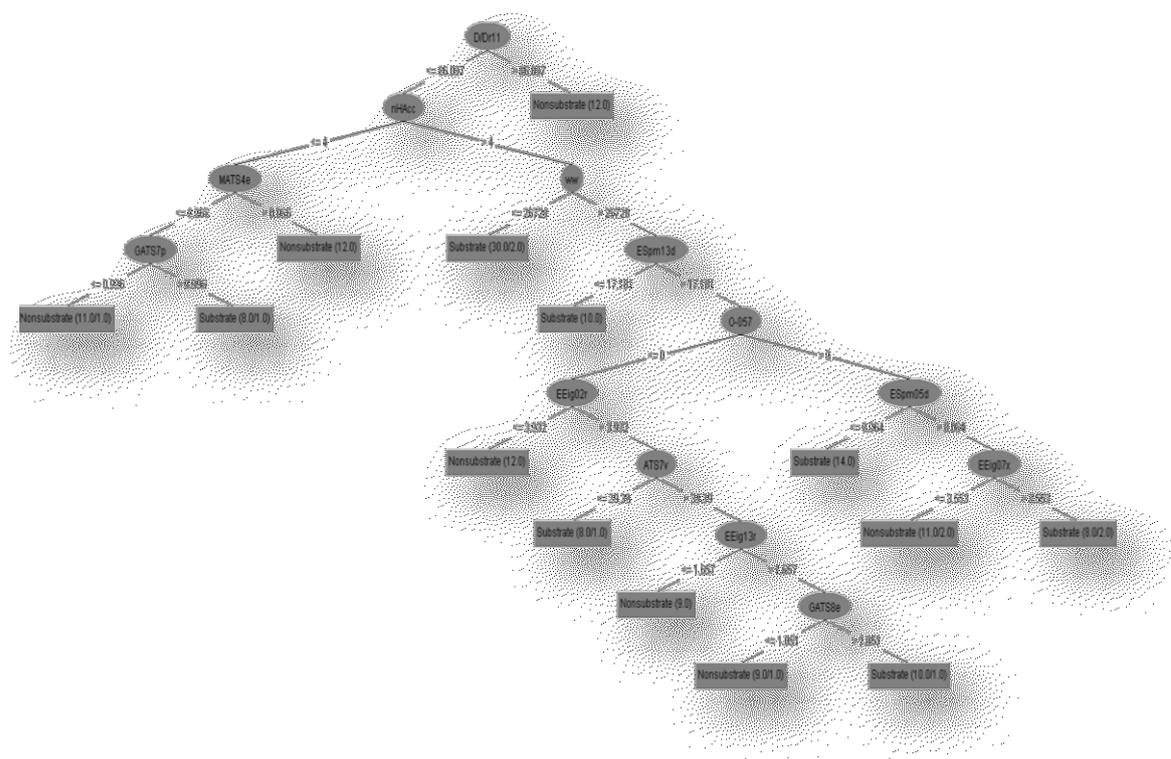
Modelo 1:



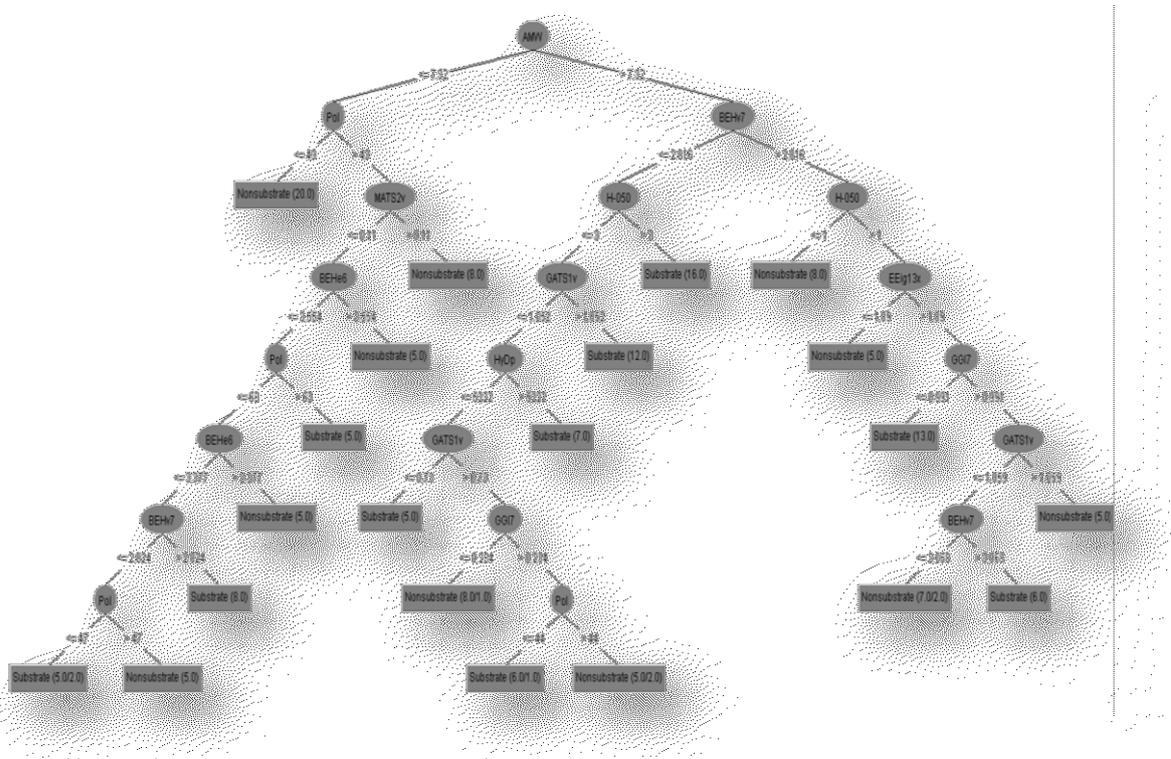
Modelo 2:



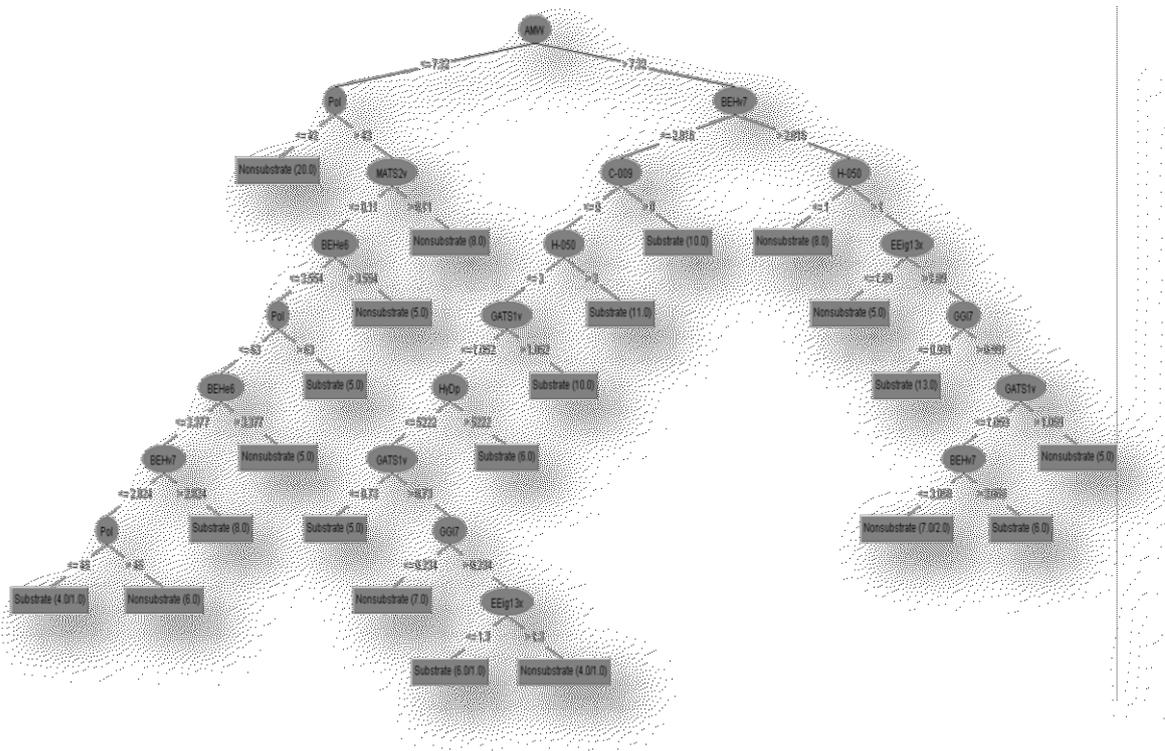
Modelo 3:



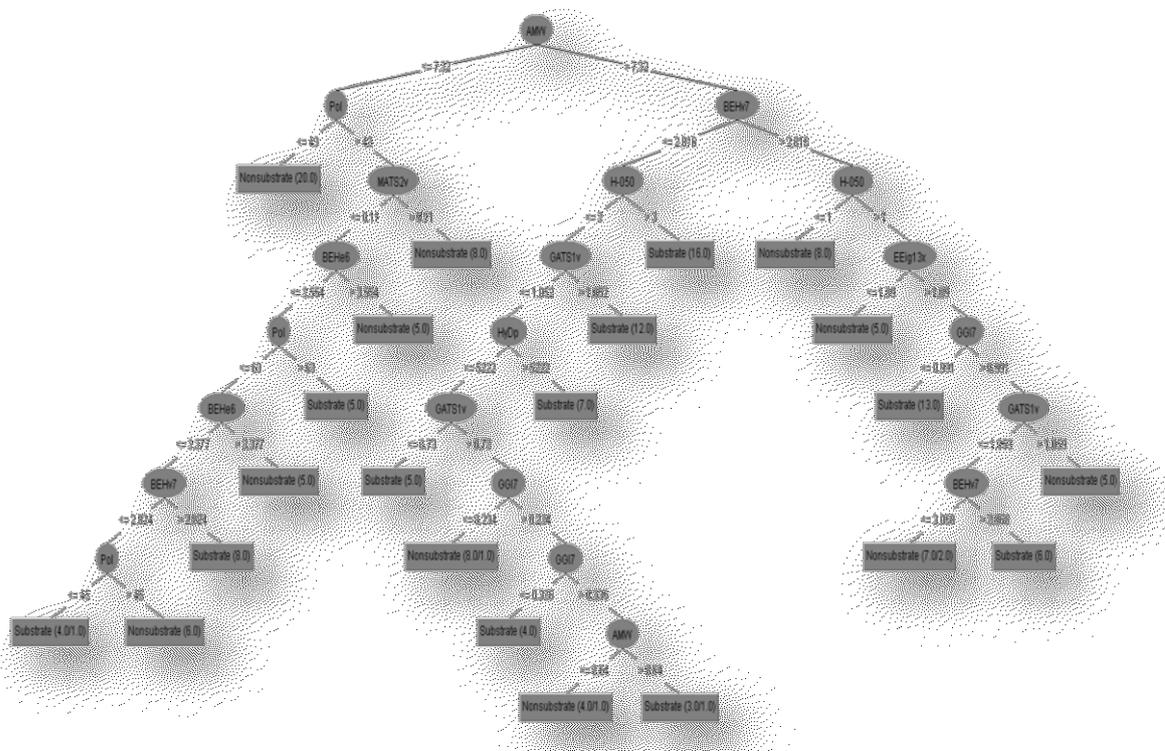
Modelo 4:



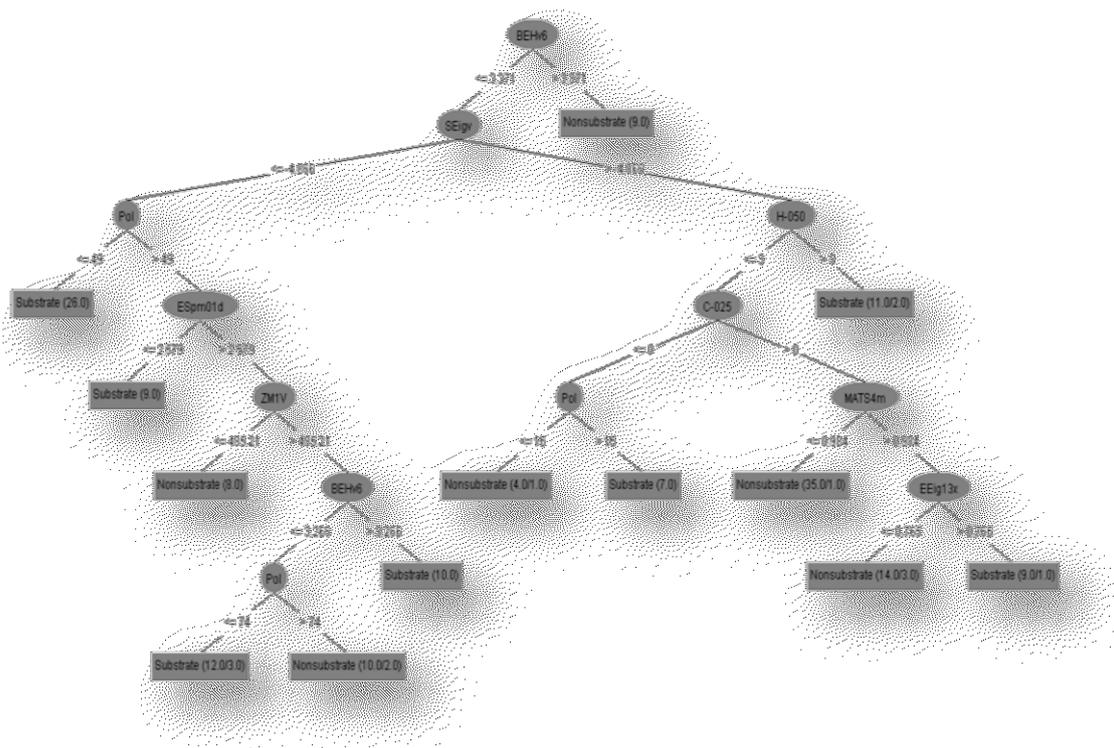
Modelo 5:



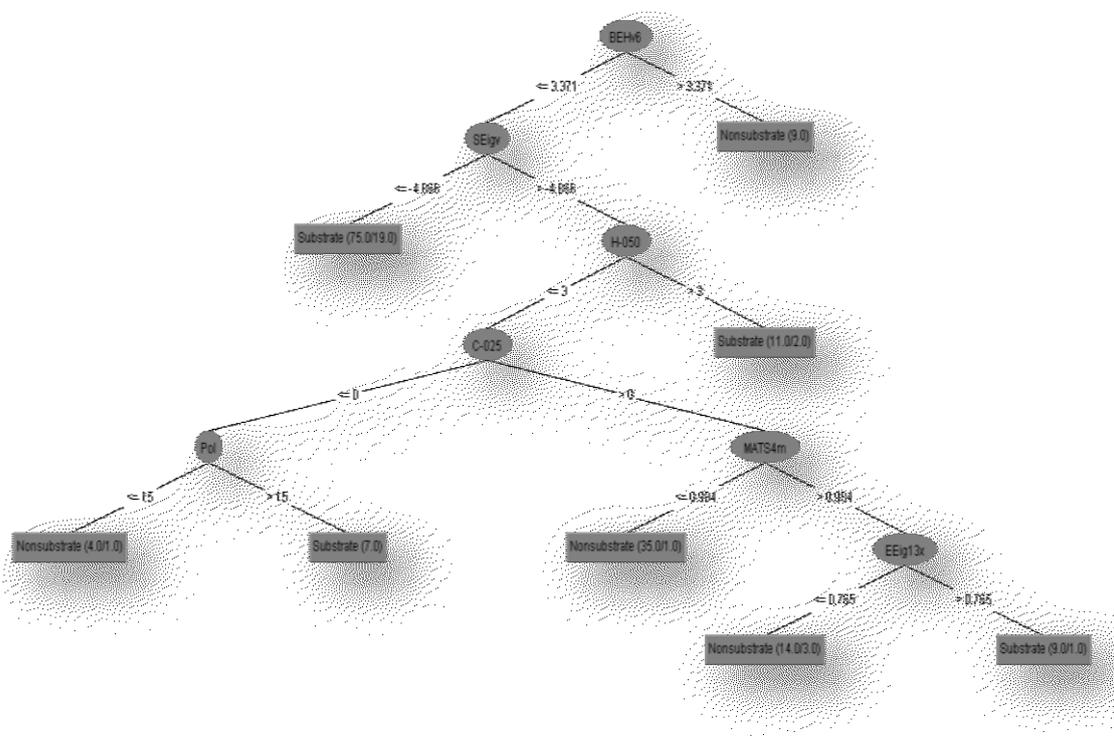
Modelo 6:



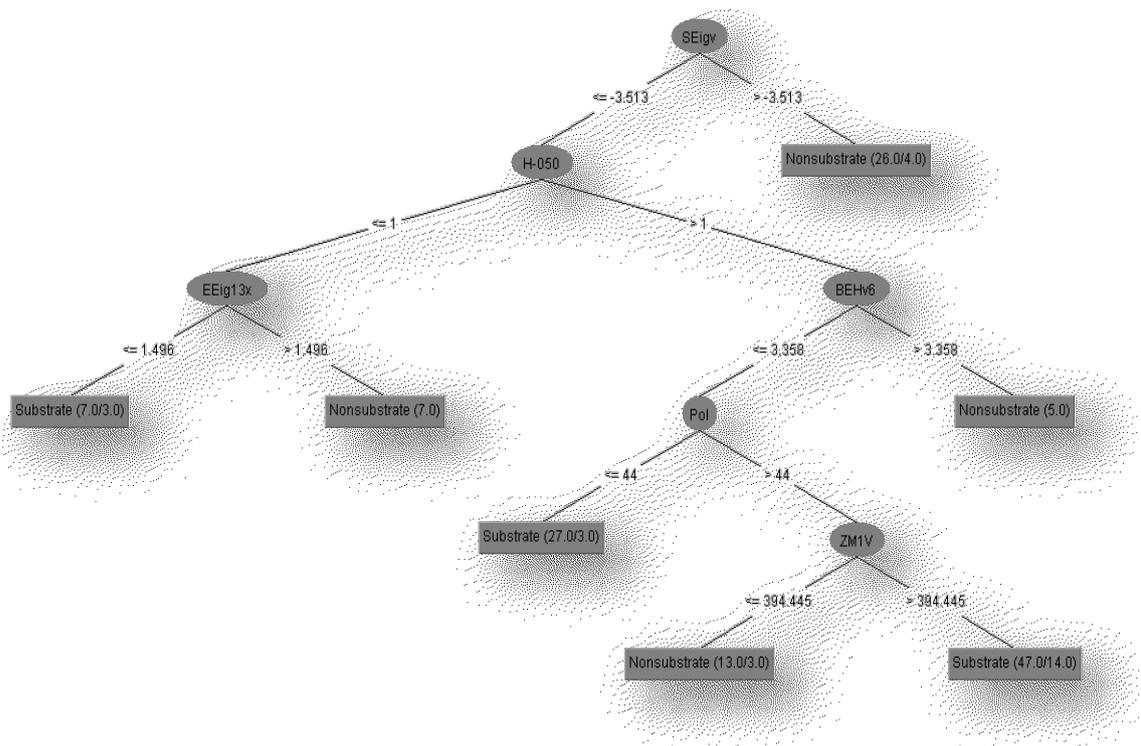
Modelo 7:



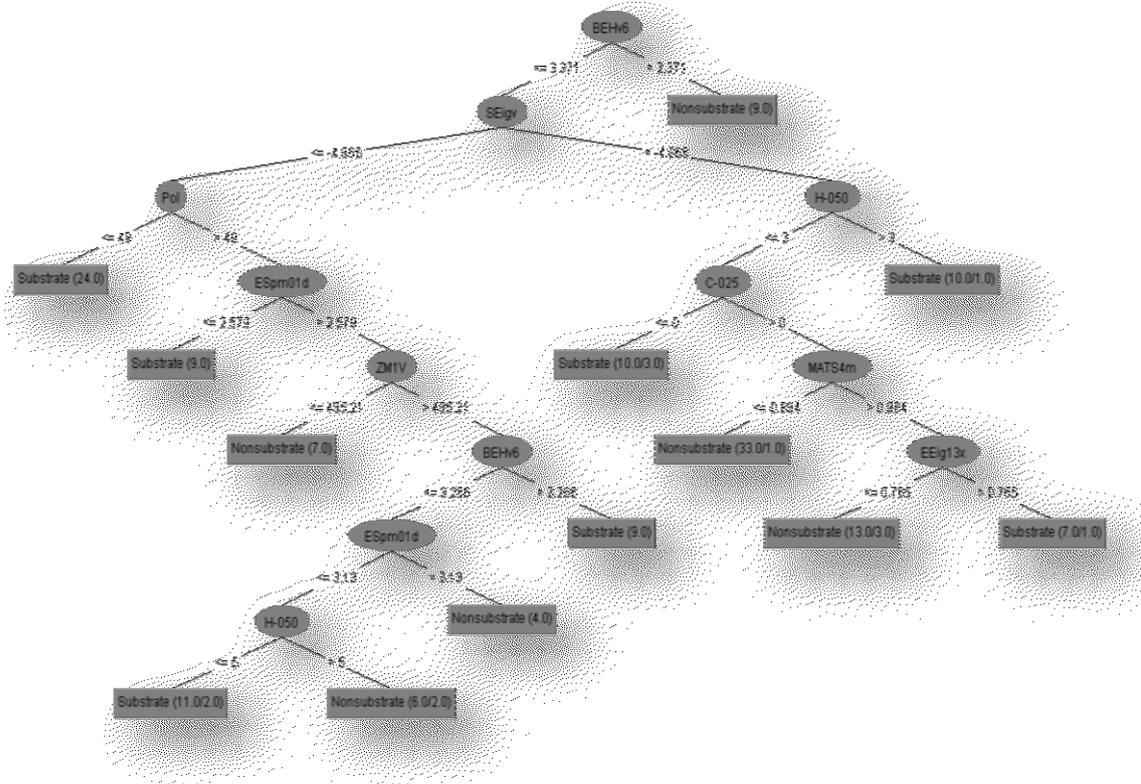
Modelo 8:



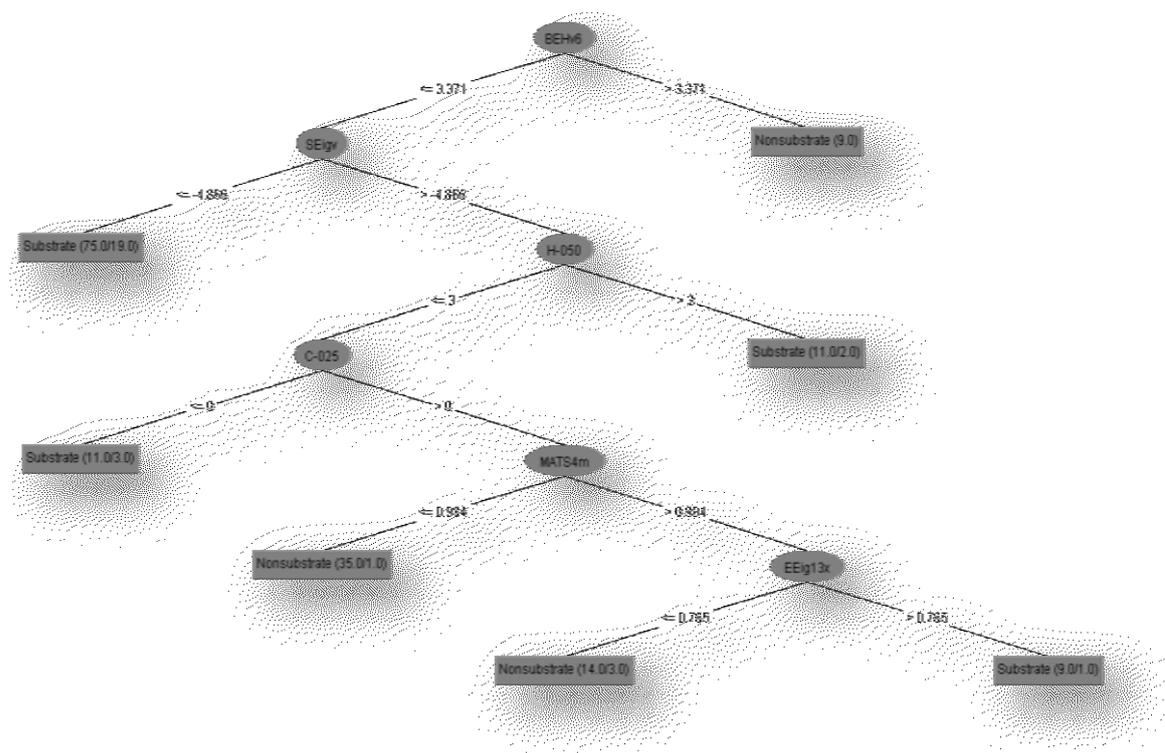
Modelo 9:



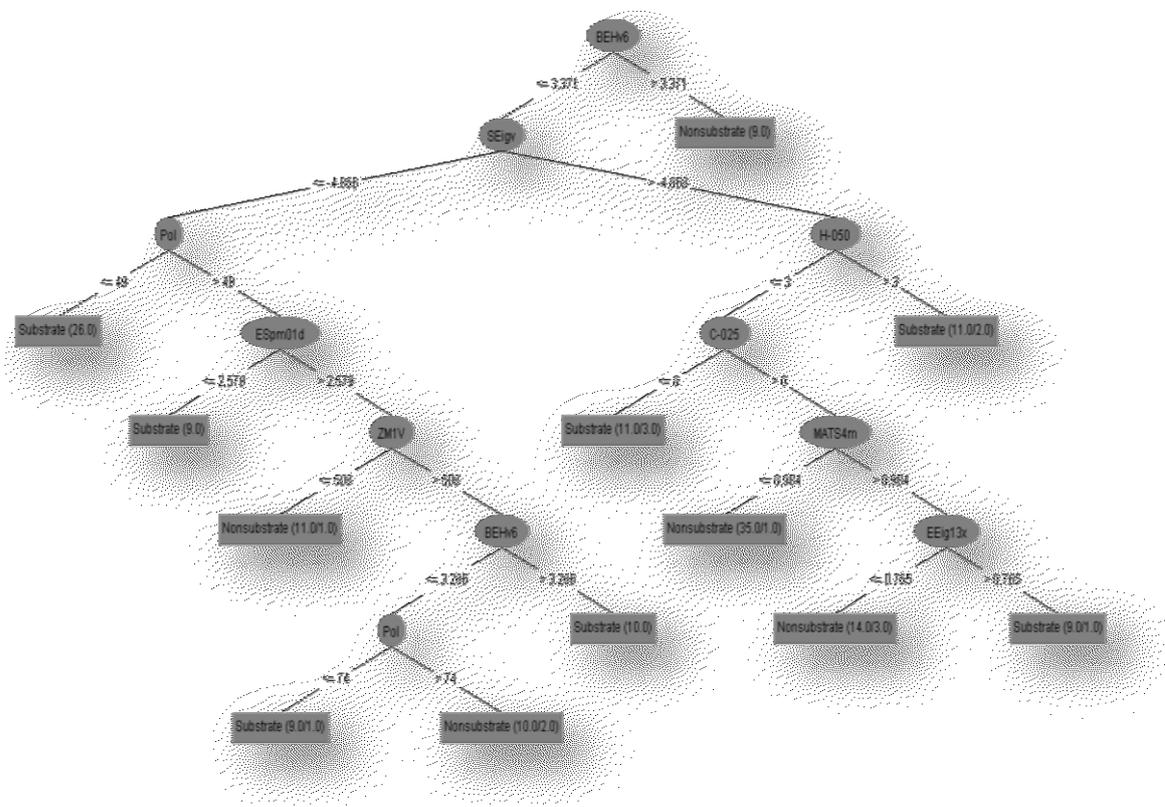
Modelo 10:



Modelo 11:



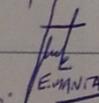
Modelo 12:

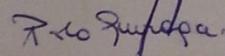


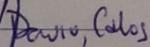
En el día de la fecha se reúne el jurado designado por el Honorable Consejo Directivo de la Facultad de Ciencias Exactas de la UNCS (Expte. 400-009649/11-005) constituido por los D<sup>es</sup> Eduardo Manta, Carlos Davio y Pablo Queropa. Para Walter <sup>atom</sup> de Tesis "Topología molecular aplicada al reconocimiento de Abstractos de la Proteína de Mama (BCRP)" bajo la Realización de la Farmacéutica Helma-Edith Santner, bajo la dirección Prof. D<sup>e</sup> Alon Talivi para optar por el grado Académico de Doctor en la "Facultad de Ciencias Exactas" área Ciencias Farmacéuticas - departamento de ciencias biológicas. Habiendo analizado el trabajo presentado por el doctorando el jurado considera que: este trabajo de Tesis aporta novedoso enfoque en el campo de la Química Medicinal para el diseño racional de nuevas estrategias terapéuticas complejas para patologías complejas como la epilepsia y el cáncer. Las metodologías utilizadas y desarrolladas para la construcción de los objetivos propuestos resultan modernas y adecuadas para lograr los propósitos pretendidos. La presentación oral por parte del doctorando fue excelente y respondió por completo las preguntas del jurado, demostrando un conocimiento profundo del tema.

Por todo lo expuesto el jurado decide Aprobar el presente trabajo y Calificarlo con 10 (Diez) Sobresaliente.

La Plata, 18 de marzo de 2016.

  
E.M.M.T.A

  
Pablo Queropa

  
Eduardo Manta