



Departamento de Ciencias Biológicas

Facultad de Ciencias Exactas

Universidad Nacional de La Plata

Screening Virtual basado en Topología Molecular: una
Nueva Metodología en el Descubrimiento de Agentes
Anticonvulsivos.



Lic. Alan Talevi

Directores: Dres. Eduardo Castro y Luis Bruno-Blanch

Año 2007

“Míralo a Euclides, por ejemplo. Puede saber exactamente la tensión que la cuerda del arco de este soldado ejerce sobre la flecha. Mide el largo del arco, calcula la distancia que hay entre las puntas del arco y la madera antes de curvarse. Pero Euclides jamás ha disparado una flecha. El soldado no es capaz de calcularlo en un papel pero sí ha disparado muchas veces muchas flechas y sus manos y sus antebrazos lo saben. La inteligencia del soldado está en sus manos y no en su mente (...). El día que Euclides quiera aprender a disparar una flecha o que este soldado se pregunte por qué sale disparada esta flecha estaremos en el camino.

Extracto de “El Depredador” , D. Sáenz

A mis abuelos Rosa y José

Agradecimientos

A mis directores, Luis y Eduardo, por su confianza y consejos.

A mis padres.

A mis compañeros de trabajo: Juan, Luciana P., Luciana G., Rubén, Ivana, Gisela, Valentina, Julián y Mariana.

A Karina, Alejandro, Mariano, Victoria, Franco, Esperanza, Agustín, Rodrigo, Kitty, Paula, Pablo.

A Carolina, por su gran ayuda.

Al CONICET y a la Facultad de Ciencias Exactas, por los medios para realizar esta tesis.

Indice

Listado de abreviaturas utilizadas en el presente trabajo	Pág. 6
Introducción	Pág. 7
<i>Evolución del método científico (Pág. 10) - Evolución de la Química Medicinal: Desde la Medicina Tradicional y la Serendipicidad al Diseño y la Búsqueda Racional de Nuevos Fármacos (Pág. 16)</i>	
Capítulo 1 - Epilepsia	Pág. 28
<i>Definición (Pág. 28) - Causas (Pág. 31) - (Dificultades en su Diagnóstico (Pág. 33) - Dificultades en su tratamiento (Pág. 35)</i>	
Capítulo 2 - Desarrollo de nuevos fármacos	Pág. 47
<i>Métodos clásicos y modernos (Pág. 47) - Relaciones Cuantitativas Estructura-Propiedad y Estructura-Actividad (Pág. 56) - Clasificación de las aproximaciones QSAR (Pág. 64) - Etapas generales de un estudio QSAR (Pág. 72)</i>	
Capítulo 3 - Descriptores utilizados en la presente tesis	Pág. 93
<i>Teoría de grafos (Pág. 93) - Matrices, polinomios y espectros derivados del grafo. Definición de algunos índices topológicos representativos (Pág. 98)</i>	
Capítulo 4 - Screening Virtual	Pág. 113
<i>Definición. ¿Por qué utilizar Screening Virtual? (Pág. 113) - Clasificación (Pág. 116) - Perspectivas (Pág. 12^o) - Aproximaciones utilizadas en el presente trabajo (Pág. 123)</i>	
Capítulo 5 - Función discriminante	Pág. 127
<i>Fundamentos del uso de la función discriminante (Pág. 127) - Pasos del modelado. Aplicación en Screening Virtual (Pág. 131)</i>	
Capítulo 6 - Ensayos biológicos	Pág. 158
Capítulo 7 - Filtros ADME	Pág. 166
<i>Filtros ADME generales (Pág. 166) - Filtros ADME/Tox particulares (Pág. 167)</i>	
Capítulo 8 - Otras aplicaciones	Pág. 186
<i>Screening Virtual como herramienta de búsqueda racional de reactivos químicos (Pág. 186) - Metodologías integradas de Screening Virtual 2D y 3D (Pág. 192) - Screening Virtual por similitud específica (Pág. 203)</i>	
Sumario y conclusiones	Pág. 216

Lista de abreviaturas utilizadas en el presente trabajo.

ADME - Absorción, Distribución, Metabolización, Excreción

ALD - Análisis Lineal Discriminante

BBB - *Blood Brain Barrier* - Barrera Hematoencefálica

DDF - Diagrama de Distribución Farmacológica

df - *Discriminant function* - Función discriminante

GA - *Genetic Algorithm* - Algoritmo genético

HTS - *High throughput screening* - screening de alta performance

LGO - *Leave Group Out* - Dejar un grupo afuera

LOO - *Leave One Out* - Dejar uno afuera

MLR - *Multiple linear regression* - Regresión Lineal Múltiple

NN - *Neural network* - Redes neuronales

QSAR - *Quantitative Structure Activity Relationship* - Relaciones cuantitativas estructura-actividad

QSPR - *Quantitative Structure Property Relationship* - Relaciones cuantitativas estructura-propiedad

REA - Relaciones estructura actividad

SNC - Sistema Nervioso Central

SV - Screening Virtual

INTRODUCCIÓN



Paralelismos entre el Desarrollo Histórico del Método Científico y la Evolución de la Química Medicinal.
De la Intuición a la Razón, y a la Intuición nuevamente.

PROMETEO: (...) Después descubrí también para ellos la ciencia del número, la más excelsa de todas, y las uniones de las letras, memoria de todo, laboriosa madre de las Musas. (...)

CORIFEO. Padeces un castigo indigno; privado de razón divagas, y como un mal médico que a su vez ha enfermado, te desanimas y no puedes encontrar para tí mismo los remedios curativos.

PROMETEO. Escucha el resto y te sorprenderás más: las artes y recursos que ideé. Lo más importante: si uno caía enfermo, no había ninguna defensa, ni alimento, ni unción, ni pócima, sino que faltos de medicinas morían, hasta que les enseñé las mezclas de remedios clementes con los que ahuyentan todas las enfermedades.

Fragmento de "Prometeo encadenado", Esquilo, S. V a.C.

"Una cosa es medir con un hilo la circunferencia de un círculo, llevar este hilo sobre el diámetro y constatar que pasa por ahí tres veces y pico; otra cosa es tratar de encontrar la relación de la circunferencia y del diámetro considerada como el límite común de los perímetros de los polígonos escritos e inscritos. A partir de entonces se trata de comprender en qué consiste este paso a la demostración (...) Pues esto es lo que se crea en Grecia por primera vez"

Fragmento de "Lo que hace a Grecia - De Homero a Heráclito", C. Castoriadis,



Evolución del Método Científico

En los principios de la civilización humana, el Conocimiento era concebido como una manifestación del aspecto divino o espiritual inherente a la naturaleza ambivalente (materia-espíritu) del hombre. Algunos de los mitos más antiguos de la humanidad se refieren justamente a una figura mitológica que actúa como intermediario entre los dioses y el hombre, transmitiendo a éste los dones propios de aquellos. Podemos mencionar como ejemplos el mito griego de Prometeo¹ (dios griego que roba a los dioses el conocimiento para regalárselo a la especie humana) y el mito egipcio de Thoth (deidad egipcia de la magia, inventor de la ciencia de los números, la astrología, la astronomía, la medicina, la escritura y la oratoria, entre otras)².



Fig. 1.1 - Prometeo, por Elsie Russell. El mito griego refiere la componente inexplicable o irracional en el desarrollo del conocimiento humano

Dentro de la tradición

¹ Es interesante mencionar que Prometeo reconoce que su mayor aporte a la humanidad es precisamente la medicina, en particular los medicamentos. El castigo de los dioses por robar este conocimiento divino no es otro que enviar un águila para que cada noche le devore el hígado, órgano que, dada la inmortalidad de Prometeo, se regeneraba durante el día. En vista de la comprensión que se tiene actualmente de la fisiología hepática, resulta irónico que el castigo del dador de los medicamentos sea justamente la destrucción reiterada del órgano metabolizador y desintoxicador por excelencia.

² Puede apreciarse, en ambos mitos, que el sistema de conocimientos ideal en la antigüedad era fundamentalmente holístico en lugar del sistema reduccionista asociado a Descartes. Todas las ciencias proceden en los mitos mencionados de un mismo dios “dador”, como si se trataran de aspectos diferenciados de una única cosa.



judeo-cristiana encontramos la historia de Semjasa [1], un ángel caído que comanda una rebelión de otros doscientos ángeles, enseñando a las mujeres humanas los secretos de las hierbas curativas, la metalurgia, la orfebrería, las tinturas y los cosméticos. La historia está registrada en el Libro de Enoch, un libro apócrifo hebreo escrito alrededor del siglo II a.C.

Semejante riqueza mitológica en torno al origen de los conocimientos que el hombre antiguo aplicaba en su vida cotidiana surgía de un desequilibrio entre la curiosidad natural del hombre y la capacidad limitada del hombre antiguo para dar explicaciones. Este misterio se presentaba magnificado, seguramente, por la ausencia de una sólida tradición escrita que permitiera registrar el origen y la evolución de los saberes humanos. Paralelamente, *es posible que este hombre antiguo hubiera identificado la inspiración o el “evento afortunado” tras el descubrimiento de un conocimiento útil*, en aquella época en la que el conocimiento se constituía sobre un conjunto de observaciones empíricas preservado y enriquecido a través de las generaciones. En uno u otro caso, el origen incierto y fortuito del saber se explicaba, como todo lo inexplicable, recurriendo a los dioses y a los mitos. Como plantea Aristóteles en el siglo IV a.C. *“Ir en busca de una explicación y admirarse es reconocer que se ignora. Y así, puede decirse que el amigo de la ciencia lo es en cierta manera de los mitos, porque el asunto de los mitos es lo maravilloso”* (*Metafísica*, Libro I, parte II) [2].

Precisamente a los griegos, en un proceso que se inicia alrededor del período homérico (hacia el siglo VIII a.C.), y en particular a Aristóteles, se les atribuye la introducción de las bases del método científico que hoy conocemos. Es lo que Otto ha dado en llamar “la superación del pensamiento mágico a partir del período



homérico” [3]. Claramente, algunos aspectos de la obra aristotélica constituyen esbozos de los principios de la ciencia moderna. Así, en *Metafísica* puede leerse “muchos recuerdos de una misma cosa constituyen una experiencia. Pero la experiencia, al parecer, se asimila casi a la ciencia y al arte³. Por la experiencia progresan la ciencia y el arte en el hombre. La *experiencia*, dice Polus, y con razón, *ha creado el arte, la inexperiencia marcha a la ventura*. El arte comienza, cuando de un gran número de nociones suministradas por la experiencia, se forma una sola concepción general que se aplica a todos los casos semejantes⁴” (*Metafísica*, Libro I, parte I) [4]. Los fundamentos de la lógica aplicados al desarrollo del método científico son explorados en profundidad por el pensador griego en sus *Segundos Analíticos* (también conocidos como *Analíticos Posteriores*): “asumiendo entonces que mi tesis sobre la naturaleza del conocimiento científico es correcta, las premisas del conocimiento demostrado deben ser verdaderas, primarias, inmediatas, mejor conocidas y anteriores a la conclusión, que se relaciona con ellas como el efecto con la causa” (*Segundos Analíticos*, Libro 1, Parte II)

De la lectura de la obra aristotélica nos llaman la atención: a) el impresionante aporte del filósofo griego al **desarrollo de la lógica y su relación con el método científico**; b) **la noción de que a la ciencia le concierne lo universal** (encontrar verdades subyacentes a todo un género de objetos y no las particularidades de un objeto individual, principio que podemos relacionar con la elaboración, por parte del científico, de modelos, teorías y leyes de aplicación general) [5]; c) **una valorización del saber del filósofo o científico frente al hombre con conocimientos técnicos** y d) **la reivindicación del conocimiento racional y sistemático frente al conocimiento “accidental”**. Sobre

³ En la Grecia antigua, la distinción entre ciencia y arte no es tan clara como (ni se corresponde con) la actual.

⁴ Como se verá más adelante, es este en cierto sentido el objeto de la presente tesis.



este último ítem enumerado Castoriadis menciona [6]: *“la cuestión esencial es que en los egipcios y en los babilonios encontramos enunciados que reposan en una curiosa mezcla de cálculo primario y sobre todo de empiria. En los griegos esto se transforma en algo totalmente diferente: en teoremas demostrados rigurosamente”*. Castoriadis agrega que el segundo elemento, y el más importante, que introducen los griegos es la **creación de un espacio ideal**, un lugar virtual que no se corresponde exactamente con el espacio real, en palabras de Platón en su Carta Séptima: *“sabemos muy bien que ninguno de los círculos reales que pueden dibujarse (...) corresponden rigurosamente a la definición de círculo como lugar de todos los puntos equidistantes de otro punto”*. Lo más notable: **el espacio ideal, euclidiano, no se corresponde con la representación del espacio físico**. Mientras en el espacio euclidiano existen rectas indefinidas, a la vez que es homogéneo e isótropo, el espacio cosmológico y físico de los griegos se representaba esférico (y por lo tanto, no había rectas sino segmentos circunscriptos a la esfera) y había sí un punto privilegiado: el centro.

La contribución más destacada de Aristóteles a la ciencia ha sido esencialmente teórica; aunque inclinado a la experimentación, sus diseños experimentales eran deficientes y lo conducían habitualmente a conclusiones equivocadas (por ejemplo, que los cuerpos más pesados caen con mayor velocidad que los más livianos y que no existen fuerzas capaces de actuar a distancia). Aún los ejemplos que postula en sus obras no se corresponden completamente con sus propios enunciados teóricos [7]. En el terreno de lo pragmático, Aristóteles fue, por excelencia, un observador y un ordenador, pero la ausencia de hipótesis y experimentación adecuada en su método hacen de la ciencia aristotélica un cúmulo de observaciones, cuando no erróneas, inconclusas o de alcance limitado.



A partir del Renacimiento, y en particular desde la Ilustración, el hombre ha acudido a la razón como la herramienta fundamental para establecer un sistema ético, estético y de conocimientos [8]. La Era de la Razón se caracterizó por instaurar (o perfeccionar, teniendo en cuenta los antecedentes griegos) un sistema filosófico basado en axiomas (verdades evidentes sobre las que se construyen otros conocimientos) pero **instituyendo a la naturaleza y no a la autoridad humana como arbitro en cuestiones científicas**. Toda fuente irracional de conocimiento y sabiduría fue desterrada⁵. Se observó la necesidad de determinar experimentalmente todas las variables que afectan a un fenómeno de la naturaleza, para luego incorporarlas en una hipótesis que quedaría demostrada si y sólo si coincidiese con relativa exactitud con la observación experimental. Como representantes históricos de ésta, la llamada Revolución Científica, podemos mencionar a Copérnico, Galileo y Newton⁶, sin olvidar el significativo aporte de Descartes a través del método reduccionista.

En este contexto, la intuición, que *intuitivamente* podemos definir como un proceso mental natural en el hombre por el cual se arriba a conclusiones sin transitar o detenerse en cada uno de los eslabones de esa concatenación de axiomas sucesivos que exige la lógica, ha sido, desde la Ilustración hasta la historia reciente, progresivamente desestimada o por lo menos subestimada. R. A. Spitz, en su obra *The First Year of Life* (International Universities Press) ha señalado que *“incluso condenamos la intuición, nos burlamos de ella en el discurso científico. Esta burla, este sarcasmo, las bromas*

⁵ El poeta inglés John Keats solía acusar a Newton y al racionalismo representado por éste de haber despojado de su encanto al arcoiris, refiriéndose a los experimentos de Newton sobre el comportamiento de la luz.

⁶ Curiosamente, no sólo era Newton un hombre profundamente religioso sino que es bien conocido su profundo interés por la Magia y la Alquimia.



en tales materias traicionan nuestra incomodidad ante lo que no podemos explicar” [9].

Indagando un poco más, encontramos que la anterior definición intuitiva se aproxima bastante a una definición más académica. J.A. Arlow definió la intuición como *“la facultad de aprehensión o cognición rápida (...) involucra la habilidad para organizar silenciosamente y sin esfuerzo aparente (por ejemplo, preconscientemente) muchas diferentes observaciones hechas a través del tiempo. El proceso llega al entendimiento sin conocimiento consciente de los pasos mentales intermedios involucrados; de ahí que el conocimiento adquirido posee una súbita, inesperada, y por lo tanto sorprendente calidad. El conocimiento obtenido por las vías de la intuición requiere una validación a través de un esfuerzo cognitivo orientado y objetivo” [10].*

Psicólogos y psiquiatras coinciden en que **existen dos estilos de pensamiento bien diferenciados, el pensamiento intuitivo y el pensamiento racional** (también referidos como **sistemas cognitivos 1 y 2** o sistemas de procesamiento experiencial y racional, respectivamente) [11,12] ⁷. El sistema 1 opera naturalmente mediante el denominado *procesamiento heurístico* (atajos cognitivos para simplificar problemas complejos; el procesamiento heurístico se caracteriza por ser rápido y automático, estar orientado hacia el resultado más que hacia el proceso, **obtener conclusiones generales a partir de unos pocos casos y representar eventos bajo la forma de imágenes holísticas y dependientes del contexto**, relacionando nueva información con esquemas previos). Aunque se ha observado que el sistema 1 es proclive a alrededor de una docena de errores sistemáticos (por lo cual es

⁷ Puede observarse que la interpretación de la psicología moderna parece sustentar la diferenciación entre conocimiento experiencial y conocimiento científico que estableció Aristóteles en el siglo IV a.C.



necesario revisar y corregir sus conclusiones a través del sistema 2) [12], entre los atributos del pensamiento intuitivo figuran el establecimiento de causalidades y propensiones, la predicción de categorías y la estimación del apoyo que una serie de observaciones provee a una hipótesis (todos ellos, como sabemos, estrechamente ligados a la actividad científica).

Numerosos estudios recientes parecen reivindicar la importancia de la intuición en diversas áreas de la ciencia (entre ellas la medicina, la psicología y la economía), esgrimiendo explícita e implícitamente la hipótesis de que el entrenamiento en una disciplina científica puede acotar los errores sistemáticos del sistema 1 y que, por lo tanto, podría aceptarse formalmente a la intuición como herramienta epistemológica [13-17]. A su vez, la intuición, como sistema cognitivo basado en representaciones holísticas/integrales del objeto de estudio, seguramente jugará un papel notable en un nuevo, incipiente escenario científico que reconoce las limitaciones de la aproximación reduccionista para modelar niveles complejos de la realidad y la necesidad de equilibrar/complementar esta aproximación con un enfoque integrador y sintético [18-20].

Evolución de la Química Medicinal: Desde la Medicina Tradicional y la Serendipicidad al Diseño y la Búsqueda Racional de Nuevos Fármacos

La historia de la Química Medicinal, entendida ésta como la rama de la Química dedicada a descubrir y desarrollar nuevos agentes terapéuticos, nos revela que esta disciplina ha evolucionado desde metodologías de descubrimientos de fármacos guiadas por la intuición y el empirismo hasta las técnicas cada vez más racionales que se emplean en la actualidad [21-23]. En el primer grupo podemos mencionar, fundamentalmente, los muchos descubrimientos que la Química Medicinal debe a la **serendipicidad** [22, 24, 25]. El concepto de serendipicidad se



refiere esencialmente un descubrimiento fortuito, aunque algunos autores han agregado a la definición de serendipicidad la necesidad de un observador que posea el conocimiento y el criterio necesarios para establecer la relación que da lugar al hallazgo.⁸ Por otro lado, no hay razón para creer que el exhaustivo e impresionante conocimiento acumulado en la Medicina Tradicional no procede sino de una compilación más o menos sistemática, según el caso, de observaciones empíricas circunstanciales [26-29].

Luego del afortunado descubrimiento de la penicilina por parte de Alexander Fleming en 1929, las entonces incipientes compañías farmacéuticas se abocaron al **screening sistemático** de microorganismos en busca de nuevos antibióticos entre las décadas de 1940 y 1970. Sus esfuerzos se vieron recompensados por el descubrimiento, no sólo de nuevos antibióticos sino también de otros fármacos (por ej. levostatina, ciclosporina e ivermectina) [30].

Conforme los conocimientos de la química de los agentes medicinales se ampliaban, se comenzaron a establecer relaciones entre la estructura de un compuesto y su actividad farmacológica; estas relaciones se observaron, en un principio, al detectar rasgos químicos comunes en compuestos pertenecientes a una misma categoría terapéutica y al observar variaciones sistemáticas de la actividad en series homólogas.

⁸Esta segunda definición corresponde a la conocida máxima de Isaac Asimov que asegura que *“la suerte favorece sólo a las mentes preparadas”* (millares de hombres contemplaron la caída de una manzana antes que Newton sin postular la existencia de un campo gravitatorio, y millares también se sumergieron en el agua antes de Arquímedes sin arribar al concepto físico de empuje).

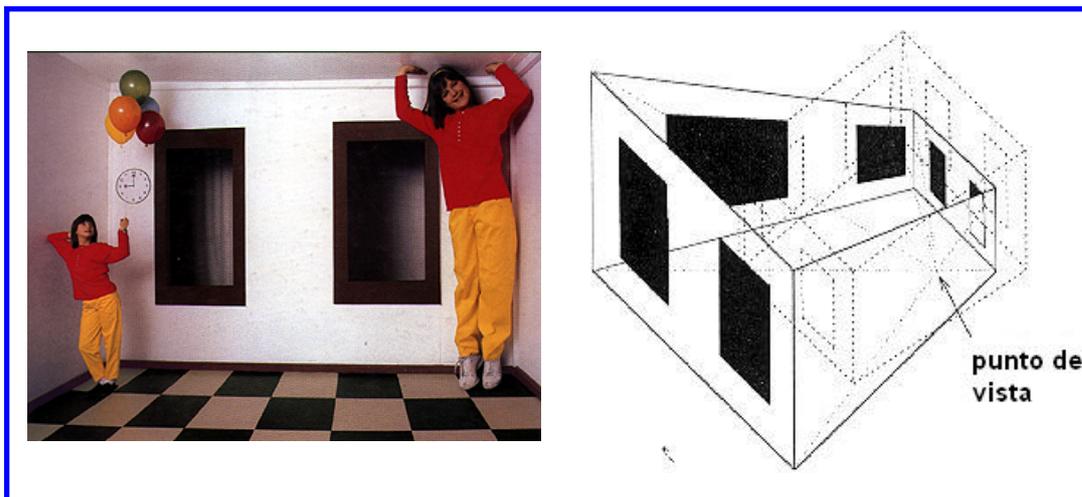


Fig 1.2 – Sobre los errores del sistema cognitivo 1 - habitación de Ames. La observación desde el punto de vista indicado en el croquis redundará en la conclusión imposible de que una persona aumenta de tamaño a medida que se desplaza desde el extremo izquierdo hacia el extremo derecho de la habitación. Esta conclusión del Sistema 1 asume que la habitación es cúbica (basado en la perspectiva y en la experiencia previa que le indica que las habitaciones suelen serlo). El sistema 2 entenderá, analíticamente, que la habitación es trapezoidal y la persona está desplazándose a través de una dirección que no es perpendicular al plano del observador.

Si bien existen algunos antecedentes aislados de este tipo de relaciones, tales como la observación por parte de Cros de que la toxicidad de los alcoholes aumenta a medida que disminuye su solubilidad en agua, en 1863, y las observaciones de Meyer y Overton respecto al aumento de la toxicidad de compuestos orgánicos con el aumento de su lipofilia, en 1890 [31,32], un punto de inflexión en la historia de la Química y la Química Medicinal puede señalarse con los estudios de Hammett respecto de la correlación entre las propiedades electrónicas de ácidos y bases orgánicos sustituidos y sus constantes de disociación, estableciendo un **orden de sustituyentes del anillo aromático del ácido benzoico según su efecto cuantitativo sobre la constante de disociación** [33]. Por otro lado, Hammett observó que un **mismo sustituyente producía efectos similares sobre otros ácidos orgánicos** (figs 1.3 y 1.4). Introducida la metodología básica de las relaciones cuantitativas estructura-propiedad (Quantitative Structure-Property Relationship - QSPR), era inminente su aplicación a



propiedades de las estructuras químicas de tanto interés como lo son las actividades biológicas, dando lugar a las relaciones cuantitativas estructura-actividad (Quantitative Structure-Activity Relationship - QSAR). Los estudios pioneros de Hansch y Fujita demostraron que la complejidad inherente a los sistemas biológicos exigía la introducción de otros descriptores complementarios a los descriptores electrónicos de Hammett, incluyendo en primera instancia descriptores para cuantificar los aspectos hidrofóbico y estérico de las moléculas estudiadas [34] y sentando las bases de los métodos de diseño y búsqueda racional de nuevos fármacos que se conocen y aplican

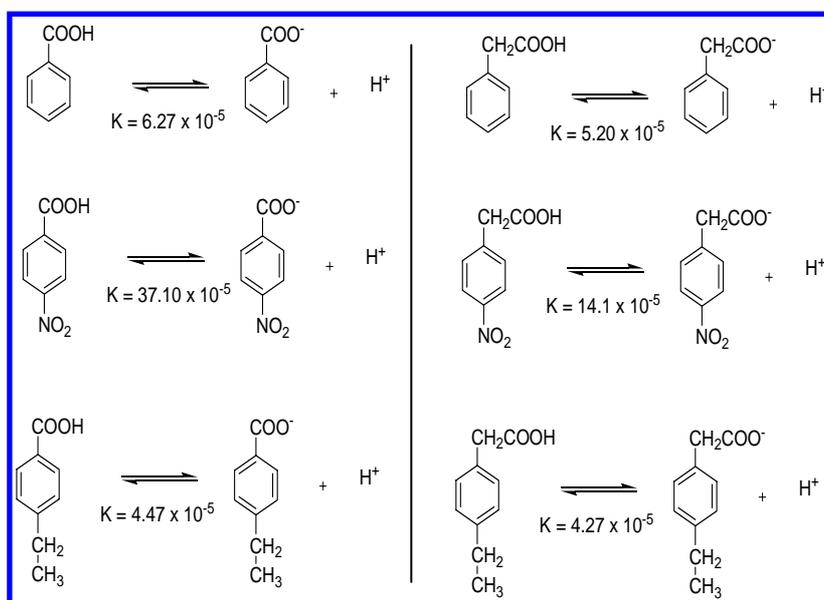


Fig. 1.3. - Hammett observó que un mismo sustituyente tenía efectos cuantitativos similares sobre la constante de disociación de distintos ácidos orgánicos; en el ejemplo se comparan los efectos de los sustituyentes nitro y etilo, aceptor y dador de electrones, respectivamente, sobre las constantes de disociación del ácido benzoico y del ácido fenílico

La tendencia actual en la Química Medicinal, distante de aquellos primeros hallazgos fortuitos y esforzados, supone racionalizar el desarrollo de nuevos agentes terapéuticos mediante relaciones estructura-actividad, y el diseño y búsqueda racional de nuevos compuestos bioactivos con posibles aplicaciones en terapéutica (metodologías que constituyen el eje fundamental de la presente

exitosamente en la actualidad, metodologías que día a día crecen en su capacidad de detectar similitudes y diferencias no evidentes entre agentes farmacoterapéuticos.



tesis) así como también mediante el modelado de los blancos moleculares y la **identificación de farmacóforos**, requisitos estructurales (electrónicos, geométricos, hidrofóbicos, estéricos) mínimos indispensables para que determinadas moléculas interaccionen efectivamente con el blanco molecular.

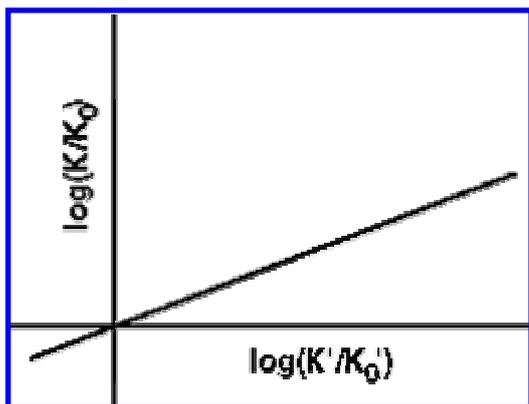


Fig 1.4 – Ejemplo de relación lineal de energía libre. K_0 y K_0' representan las constantes de disociación de los ácidos orgánicos sin sustituir (en el ejemplo de la figura tres, ácido benzoico y fenilacético) y K y K' las constantes correspondientes a los compuestos sustituidos. Los valores de la ordenada se calculan como el log de la relación entre las constantes del ácido benzoico sustituido y sin sustituir; los valores de la abscisa a partir de la misma relación para el ácido fenilacético, utilizando el mismo patrón de sustitución.

Como se ha visto, la historia de la Química Medicinal imita de alguna manera la evolución histórica de la ciencia (fig. 1.5); toda ciencia parte de conocimientos empíricos obtenidos en primera instancia en forma fortuita y luego en forma sistemática (conforme lo van permitiendo los avances tecno y epistemológicos).

Esos conocimientos, esa experiencia (en el sentido aristotélico) se acumulan y finalmente permiten la aparición de metodologías racionales. En el caso de la Química Medicinal las

herramientas que se han constituido en máximos exponentes de racionalización en la búsqueda de nuevos fármacos son las relaciones cuantitativas estructura-actividad y el modelado molecular, aplicados ambos al diseño y búsqueda de nuevos agentes terapéuticos. Sin embargo, puede observarse en la ciencia moderna una tendencia a complementar la aproximación racionalista con aproximaciones menos racionales, legitimándose la incorporación de procesos cognitivos distintos del analítico-racionalista-reduccionista a la labor científica. Castoriadis (op. cit.) dice que la magia no es en absoluto



la ausencia de leyes: “*con las magias elaboradas nos encontramos con una enorme legislación prenatal que domina o condiciona la acción de las simples regularidades naturales, pero la acción de estas leyes implica intervenciones personales del mago*”. La ciencia ha advertido que la irracionalidad y el desorden pueden ser muy productivos [35], y en lugar de descartarlos, como otrora, parece existir un movimiento que busca entender la complejidad y las leyes de estos fenómenos, para aprovecharlos. ¿La ciencia se mueve hacia un nuevo paradigma científico? Y si ese fuera el caso, ¿cómo imitará ese movimiento la Química Medicinal?

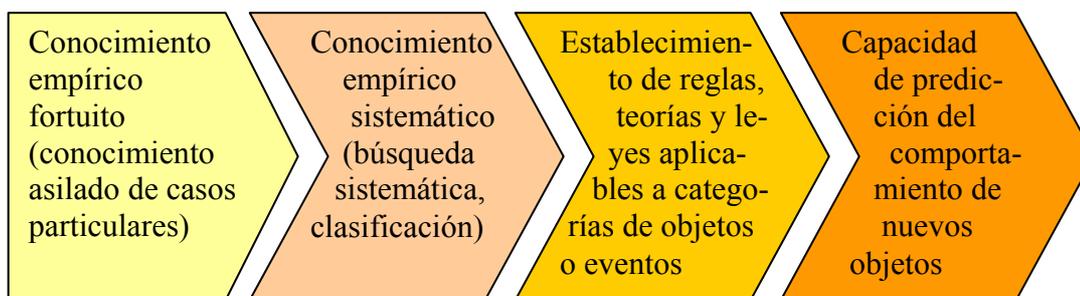


Fig. 1.5 - Esquema propuesto para explicar la evolución de una ciencia. Hacia la derecha se progresa hacia una mayor racionalización en la búsqueda de conocimiento; sin embargo, la evolución hacia estadios más avanzados no significa la extinción de estadios anteriores: los principios de todas las etapas conviven integrados. En el ámbito de la Química Medicinal, aún hoy, son comunes los reportes de hallazgos fortuitos, que coexisten con el diseño de fármacos y la búsqueda sistemática. Existen varias explicaciones para esta convivencia relativamente pacífica de aproximaciones más o menos racionales: no es ético descartar un nuevo conocimiento por su carácter fortuito, por un lado, y, por otro, el paradigma de novedad y velocidad propio de nuestra época.

La presente tesis se dividirá en dos partes. La primera de ellas, de carácter introductoria, versará sobre las motivaciones y los fundamentos del trabajo de tesis, describiendo la epilepsia y el estado actual de su terapéutica y proporcionando el marco teórico para comprender las bases y los alcances de los **estudios QSAR**, la **Topología Molecular** y las técnicas de **Screening Virtual (SV)**. En la segunda se expondrán los resultados obtenidos mediante aplicación de estas metodologías en la búsqueda de nuevos fármacos anticonvulsivos.



Pese a que el trabajo se enmarca en la aproximación racional en la búsqueda de nuevos agentes anticonvulsivos con mecanismo de acción asociado a bloqueos de canales de sodio, estoy profundamente convencido de que la razón y la intuición y lo racional e irracional no son principios mutuamente excluyentes sino complementarios en la labor científica y al accionar humano.

A modo de ejemplo de esta “contaminación” de la razón por lo irracional (no se pretende darle a contaminación una connotación negativa) en la búsqueda de agentes terapéuticos se puede considerar el **Screening de Alta Performance** (High-throughput screening - HTS), asociado a las modernas técnicas de síntesis por **Combinatoria Química**. Esta técnica, de amplio uso en la actualidad, supone ensayar, mediante sistemas robotizados, una gran cantidad de compuestos de gran diversidad química a través de un gran número de ensayos in-vitro diversos (binding, inhibición o interacción con enzimas, etc.) [36]. Puede observarse que esta aproximación no difiere en sus fundamentos del ya mencionado screening sistemático de antibióticos llevado a cabo entre 1940 y 1970 por las compañías farmacéuticas, por cuanto aquel mismo enfoque se ha asociado a las modernas tecnologías de síntesis y evaluación farmacológica permitiendo el screening de una gran cantidad de compuestos, en un tiempo mínimo y racionalizando la cantidad de muestra sintetizada y empleada en los ensayos. Sin embargo, la esencia de esta estrategia es la empiria, sistematizando, con el fin de lograr resultados positivos en un tiempo pequeño, el proceso que la Medicina Tradicional llevaba a cabo antiguamente mediante hallazgos fortuitos que se acumulaban a través de las generaciones. Se trata de los mismos principios adaptados, tecnología mediante, al que parece ser el paradigma moderno: novedad y velocidad. Advirtiendo esta característica (a punto tal que el HTS ha sido denominado “engineered



serendipity - serendipicidad dirigida”), aproximaciones novedosas proponen que la componente irracional de esta metodología puede minimizarse complementándola con una preselección racional (a través del modelado y del SV) de las estructuras que serán sometidas a HTS orientado a objetivos terapéuticos determinados, utilizando además filtros virtuales descartar compuestos con características químicas indeseadas (por ej. alta reactividad, presencia de toxicóforos) [36, 37].

Un segundo ejemplo lo constituyen las técnicas de SV. Aún cuando son un exponente de máxima racionalización en el desarrollo y descubrimiento de nuevos fármacos, un estudio de SV aplicado a grandes colecciones virtuales de compuestos (típicamente del orden de 10^5 a 10^6 estructuras químicas virtuales) suele seleccionar numerosos compuestos potencialmente activos. La evaluación biológica exhaustiva de esos compuestos no siempre es factible, por motivos económicos e incluso en algún punto por razones bioéticas, lo que implica que el investigador suele realizar una segunda selección en base a criterios diversos (novedad estructural, criterios químicos y también, en última instancia, mediante una selección intuitiva) en los que se ponen en juego ambos sistemas cognitivos, 1 y 2, en forma simultánea e integrada.

En síntesis, podemos intentar minimizar la componente irracional del estudio científico (el desarrollo de fármacos incluido) pero puede que no sea posible ni producente aniquilarla.

Es bien conocida la anécdota del descubrimiento de la penicilina. Se dice que Fleming era extremadamente desordenado. Un día, accidentalmente, estornudó sobre un cultivo de bacterias, encontrando luego un halo de inhibición en el lugar en que había



estornudado, lo que lo llevó al descubrimiento de la lisozima presente en las secreciones nasales y su moderado poder antibiótico. Más adelante, luego de regresar de un período de descanso, encontró un segundo cultivo contaminado con una cepa del género *Penicillium*. Alrededor del moho también se había formado un halo de inhibición. Fleming relacionó este hecho con el anterior y eso lo condujo al descubrimiento de la penicilina. Años más tarde, mientras visitaba un laboratorio muy limpio y ordenado, otro científico exclamó:

-¡Qué maravillosos descubrimientos podría haber hecho Fleming aquí!

A lo que Fleming contestó:

- Seguro que no el de la penicilina.

Referencias

1. Schummer, J. The notion of nature in Chemistry. *Stud. Hist. Phil. Sci.* **2003**, 34, págs. 705-736
2. Aristóteles. Metafísica. Traducción de Patricio de Ascárate. Biblioteca Virtual Miguel de Cervantes, Alicante, **1999**. <http://www.cervantesvirtual.com> (basada en la edición de Miguel Candel, Espasa Calpe, Madrid, 1997)
3. Otto, W.F. Los dioses de Grecia. Siruela, Madrid, **2003**.
4. Aristóteles. Posterior Analytics. Traducción de G.R.G. Mure. Internet Classics Archive. <http://classics.mit.edu>
5. Marcos A. Towards a science of the individual: the Aristotelian search for scientific knowledge of individual entities. *Stud. Hist. Phil. Sci.* **2004**, 35, págs. 73-89
6. Castoriadis, C. Lo que hace a Grecia. 1. De Homero a Heráclito. Seminarios 1982-1983. La Creación Humana. Fondo de Cultura Económica, Buenos Aires, **2006**, págs. 79-100.
7. Oriol Salgado, M. La estructura del silogismo práctico en Aristóteles. *Revista de Filosofía*. **2004**, 29, págs. 53-75
8. Oliveras, E. Estética, la cuestión del arte. Ariel, Buenos Aires, **2005**, págs. 23-26



9. Piha, H. Intuition: a bridge to the coenesthetic world of experience. *J. Am. Psychoanal. Assoc.* **2005**, 53, págs. 23-49
10. Arlow, J.A. The genesis of interpretation. *J. Am. Psychoanal. Assoc.* **1979**, 27 (suppl.), págs. 193-206
11. Shiloh, S.; Salton, E.; Sharabi, D. Individual differences in rational and intuitive thinking styles as predictors of heuristic responses and framing effects. *Pers. Ind. Diff.* **2002**, 32, págs. 415-429
12. Kahnemann, D. A perspective on judgment and choice. Mapping Bounded Rationality. *Am. Psychol.* **2003**, 58, págs. 697-720
13. Greenhalgh, T. Intuition and evidence - uneasy bedfellows? *Brit. J. Gen. Pract.* **2002**, 52, págs. 395-400
14. Lichtenstein, A. Integrating Intuition and Reasoning. How Balint groups can help medical decision making. *Aust. Fam. Phys.* **2006**, 35, págs. 987-989
15. Ruth-Sadh, L.A.; Hendy, H.M. Predictors of novice nurses' use of intuition to guide patient care decisions. *J. Nurs. Educ.* **2005**, 44, págs. 450-458
16. Smith, A.J.; Thurkettle M.A.; de la Cruz, F.A. Use of intuition by nursing students: instrument development and testing. *J. Adv. Nurs.* **2004**, 47, págs. 614-622
17. Leybourne, S.; Sadler-Smith, E. The role of intuition and improvisation in project management. *Int. J. Proj. Mgmt.* **2006**, 24, págs. 483-492
18. Emblemsvåg, J.; Bras, B. Process thinking - a new paradigm for science and engineering. *Future.* **2000**, 32, págs. 635-654
19. Skurvydas, A. New methodology in biomedical science: methodological errors in classical science. *Medicina (Kaunas).* **2005**, 41, págs. 7-16
20. Andersen, H. The history of reductionism versus holistic approaches to scientific research. *Endeavour.* **2001**, 25, págs. 153-156
21. Pauwels, R. Aspects of successful drug discovery and development. *Antivir. Res.* **2006**, 71, págs. 77-89
22. Kubinyi, H. Chance favors the prepared mind - from serendipity to rational drug research. *J. Recept. Signal Transduct. Res.* **1999**, 19, págs. 15-39
23. Markman, M.; Peereboom, D.M. From serendipity to design: the evolution of drug development in oncology. *Cleve. Clin. J. Med.* **1997**, 64, págs. 155-163
24. Dousset, J.C. Serendipity and pharmaceutical research. *Rev. Hist. Pharm.* **2003**, 51, págs. 191-196
25. Ban, T.A. The role of serendipity in drug discovery. *Dialogues Clin. Neurosci.* **2006**, 8, págs. 335-344



26. Patwardhan, B. Ethnopharmacology and drug discovery. *J. Ethnopharmacol.* **2005**, 100, págs. 50-52
27. Bussmann, R.W.; Sharon, D. Traditional medicinal plant use in Northern Peru: tracking two thousand years of healing culture. *J. Ethnobiol. Ethnomed.* **2006**, 2:47
28. Feng, Y.; Wu, Z.; X. Zhou; Z. Zhou; W. Fan. Knowledge discovery in traditional Chinese medicine: State of the art and perspectives. *Artif. Intell. Med.* **2006**, 38, págs. 219-236
29. Ohtake, N.; Nakai, Y.; Yamamoto, M.; Sakakibara, I.; Takeda, S.; Amagaya, S.; Aburada, M. Separation and isolation methods for analysis of the active principles of Sho-saiko-to (SST) oriental medicine. *J. Chromatogr. B.* **2004**, 812, págs. 135-148
30. Walker, M.J.A.; Barrett, T.; Guppy, L.J. Functional pharmacology: the drug discovery bottleneck? *Drug Discov. Today: Targets.* **2004**, 3, 208-215
31. Borman, S. New QSAR techniques eyed for environmental assessments. *Chem. Eng. News.* **1990**, 68, págs. 20-23
32. Lipnick, R.L. Charles Ernest Overton: narcosis studies and a contribution to general pharmacology. *Trends Pharmacol. Sci.* **1986**, 7, págs. 161-164
33. Hansch, C.; Leo, A.; Taft, R.W. A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.* **1991**, 91, págs. 165-195
34. Hansch, C. A Quantitative approach to biochemical structure-activity relationships. *Acct. Chem. Res.* **1969**, 2, págs. 232-239
35. Abrahamson, E.A.; Freedman, D.H. A perfect mess. The hidden benefits of disorder. Little, Brown and Company, London, **2007**
36. Macarron, R. Critical review of the role of HTS in drug discovery. *Drug Discov. Today.* **2006**, 11, págs. 277-279
37. Good, A.C.; Krystek, S.R.; Mason, J.S. High-throughput and virtual screening: core lead discovery technologies move towards integration. *Drug Discov. Today,* **2000**, 5, págs. 561-569

PARTE I

MARCO TEÓRICO – CONCEPTUAL



Capítulo 1. Epilepsia.

- Definición.
- Causas.
- Dificultades en su Diagnóstico y Tratamiento.

Definición

La epilepsia es el desorden crónico del sistema nervioso central (SNC) de mayor prevalencia: alrededor del 1-2% de la población mundial padece de epilepsia, y el porcentaje de la población afectado casi se duplica en grupos poblacionales específicos tales como niños, adolescentes y ancianos [1]. El 80% de los pacientes epilépticos es originario de países subdesarrollados.

Si bien tradicionalmente se ha caracterizado a la epilepsia por la ocurrencia de crisis o episodios epilépticos no provocados y reiterados (dos o más), revisiones recientes de la definición de epilepsia por parte de organizaciones internacionales (International League Against Epilepsy e International Bureau for Epilepsy) proponen que no se considere la reiteración del episodio epiléptico como criterio de diagnóstico excluyente, sino que se incorpore como criterio de diagnóstico **la ocurrencia de un único episodio y la presencia de daño o alteración permanente del cerebro que aumente la probabilidad de nuevos episodios**¹ [2].

Los síntomas de la epilepsia se manifiestan como consecuencia de una descarga anormal de neurotransmisores, excesiva y sincrónica, a

¹ Este segundo criterio de diagnóstico no es caprichoso, por cuanto las probabilidades de remisión de la enfermedad disminuyen conforme se producen nuevos episodios; es decir, las chances de recuperación del paciente epiléptico aumentan con un rápido diagnóstico e inicio del tratamiento.



nivel del SNC. Las características del episodio epiléptico pueden ser muy variadas, yendo desde un simple lapso en la atención hasta convulsiones severas. La diversidad de la sintomatología, como veremos más adelante, dificulta en muchos casos su diagnóstico.

La Tabla 2.1 presenta una clasificación general de los desórdenes epilépticos y las características clínicas de los tipos más comunes [3].

Tabla 2.1 – Una de las posibles clasificaciones de los desórdenes epilépticos más comunes (se han identificado más de 40 síndromes epilépticos distintos).

Tipo de crisis epilépticas		Características Clínicas
Parciales: Se originan en una porción del cerebro	Parciales simples (focales)	No hay pérdida de conciencia; los síntomas pueden ser motores, sensoriales, autonómicos o psíquicos, dependiendo del foco de la descarga neuronal
	Parciales complejas (del lóbulo temporal o psicomotoras)	El episodio puede comenzar sin aviso o con signos o síntomas motores, sensoriales, autonómicos o psíquicos. Hay pérdida de conciencia. Pueden ocurrir automatismos; al episodio lo sobreviene un periodo de confusión
	Parciales secundariamente generalizadas (tónico-clónicas o grand mal)	El episodio puede comenzar con signos o síntomas motores, sensoriales, autonómicos o psíquicos. Hay pérdida de conciencia, con incremento tónico del tono muscular seguido de contracciones rítmicas (mioclónicas). Luego de la convulsión el paciente está comatoso y se recupera lentamente; el paciente puede morderse la lengua o sufrir incontinencia
Generalizadas: La descarga	De ausencia (petit mal)	El episodio comienza rápidamente y consiste en un breve lapso de conciencia (en promedio, diez



eléctrica que conduce al episodio involucra todo el cerebro	Primariamente generalizadas tónico-clónicas (grand mal)	segundos) con recuperación rápida; puede haber decremento o incremento del tono muscular, automatismos y movimientos clónicos suaves. El episodio puede desencadenarse por hiperventilación; la edad de la primera crisis va generalmente de los 3 a los 20 años Ocurre pérdida de conciencia sin aviso o precedida por contracciones mioclónicas; las características clínicas son similares a las de los episodios parciales secundariamente generalizados
---	---	---

La enfermedad está asociada a un aumento del riesgo de mortalidad debido a:

- una enfermedad cerebral subyacente, como un tumor o una infección;
- la ocurrencia del episodio en circunstancias peligrosas en las que el paciente pueda ahogarse, quemarse, o sufrir heridas en la cabeza;
- *status epilepticus* (estado en el que el paciente sufre crisis epilépticas de duración mayor a 30 minutos o bien crisis de menor duración, con alta frecuencia, sin recuperar la conciencia entre episodios subsecuentes, pudiendo producirse daño cerebral o muerte [4]);
- muerte súbita debida a paro respiratorio o cardio-respiratorio durante un episodio;
- suicidio.

El aumento de riesgo de suicidio en los pacientes epilépticos no es extraño: en todas partes del mundo, aún en los países



desarrollados, las personas que padecen de esta enfermedad son socialmente estigmatizadas [5]. A la imposibilidad que el paciente epiléptico enfrenta para desarrollar actividades comunes tales como conducir un vehículo, nadar u operar maquinaria, se suman sus dificultades para conseguir trabajo y problemas de discriminación. Aún en países industrializados como Alemania, Italia y Estados Unidos, sólo un 40 a 60% de la población epiléptica en edad productiva consigue empleo; entre 15 a 20% se haya desocupada y un 20% se jubila en forma prematura. Usualmente, el paciente consigue un trabajo por debajo de las expectativas de una persona sana con capacitación similar. En China la epilepsia es vista como causa válida para prohibir o anular matrimonios, y, en muchos países de África y Asia (Liberia, Camerún, Nepal, India e Indonesia, entre otros) la enfermedad es popularmente asociada, aún en el presente, a posesión demoníaca y brujería.

Causas

Las crisis epilépticas son un síntoma de un desorden subyacente cuya causa puede ser [1, 3, 6-8]:

- Genética: patologías debidas a mutaciones en genes que codifican subunidades de canales iónicos voltaje-operados o ligando-operados (por ej., la epilepsia del lóbulo frontal nocturna dominante autosómica -ADNFLE- se relaciona con mutaciones en los genes que codifican las subunidades $\alpha 4$ y $\beta 2$ de los receptores de acetilcolina neuronales y la epilepsia mioclónica juvenil autosómica dominante se asocia a una mutación de la subunidad $\alpha 1$ del receptor GABA-A); también podemos incluir las epilepsias vinculadas a



condiciones neurológicas progresivas tales como encefalopatías mitocondriales y fenilcetonuria;

- Traumática: debida a lesiones en la cabeza (tanto heridas abiertas, incluidas las heridas por intervención quirúrgica, heridas cerradas con hematoma intracraneal o contusión hemorrágica);
- Metabólica: debida a hipoglucemia, anoxia cerebral, desbalance electrolítico;
- Infecciosa: como secuela de infecciones virales o bacterianas (meningitis, tuberculosis) o por parásitos (malaria, esquistosomiasis, neurocisticercosis). **Enfermedades febriles de cualquier clase** pueden disparar crisis epilépticas en niños pequeños, asociadas a incremento del riesgo de desarrollar epilepsia posteriormente;
- Maligna: por presencia de tumores cerebrales;
- Fármaco y Toxicológica: en pacientes con historia de abuso crónico de alcohol, drogas “recreacionales” como la cocaína y drogas terapéuticas como los barbituratos y las benzodiacepinas (usualmente, se desencadena en períodos de abstinencia). También en personas expuestas al plomo, el monóxido de carbono, insecticidas organofosforados y otras sustancias neurotóxicas;
- Causas vinculadas a factores de riesgo pre o perinatales, tales como encefalopatías neonatales severas con desorden motor residual o malformaciones en el cerebro;
- Enfermedades o accidentes cerebrovasculares, tales como infarto o hemorragia cerebral.

En síntesis, **cualquier factor que exponga el cerebro a daño o funcionamiento anormal puede convertirse, eventualmente, en causa de epilepsia, y la probabilidad de desarrollar la enfermedad aumenta**



cuanto mayor es el tiempo de exposición al factor de riesgo o por re-exposición al mismo. El término *epilepsia criptogénica* se utiliza para describir aquellos casos de epilepsia no asociados a un factor de riesgo conocido.

Observando la lista de factores de riesgo pueden comprenderse los motivos por los cuales la prevalencia de la enfermedad es mucho mayor en los países subdesarrollados. En los países desarrollados, las mejores condiciones sanitarias permiten extremar los cuidados peri y postnatales y durante el trabajo de parto, y mayor control de la fiebre en niños pequeños. Por otro lado, existe menor prevalencia de infecciones bacterianas y por parásitos, índices mucho menores de desnutrición e incluso una cultura de la prevención que reduce el riesgo de daño cerebral (controles periódicos de la presión sanguínea, uso del cinturón de seguridad y cascos).

Dificultades en el Diagnóstico

Las manifestaciones clínicas de la epilepsia en ocasiones pueden conducir a un diagnóstico erróneo por parte del médico, que puede o bien confundir el tipo de epilepsia que presenta el paciente (prescribiendo un tratamiento farmacológico que no es el más adecuado, lo cual puede agravar los síntomas en lugar de controlarlos) o diagnosticar una patología distinta de la epilepsia [9]. La falla en el diagnóstico es significativa, por cuanto **la reincidencia de la crisis epiléptica deviene usualmente en la aparición de epilepsia**, entendida como ya se dijo como una susceptibilidad o predisposición a padecer crisis en forma crónica.

Dos problemas comunes en el diagnóstico de epilepsia son diferenciar un episodio con convulsiones tónico-clónicas de un síncope



y distinguir entre una crisis de ausencia y una crisis parcial compleja. En ambos casos suele requerirse la observación de nuevos episodios por parte del médico para un diagnóstico exacto. En el primer caso, la convulsión tónico-clónica puede manifestarse mientras el paciente está despierto o dormido y en cualquier posición, una duración de más de un minuto, aumento del tono muscular durante el evento, incontinencia, mordedura de lengua, rubor y piel caliente y sudorosa y **electroencefalograma (EEG) anormal**; el síncope, en cambio, ocurre sólo cuando el paciente se haya erecto, se caracteriza por una duración similar a 10 segundos, flacidez muscular, palidez, y piel fría y húmeda. En el segundo caso, tanto la crisis de ausencia como las crisis parciales complejas pueden caracterizarse por una pérdida de conciencia, diferenciándose esencialmente por la duración de la crisis y el tiempo de recuperación y **por las características del EEG**. Se aprecia que el EEG puede ser un elemento de diagnóstico fundamental. Esta herramienta permite confirmar actividad eléctrica anormal en el SNC, reunir información sobre el tipo de epilepsia presente y determinar el foco de la crisis. En el 50% de los pacientes es necesario realizar por lo menos dos EEG para un diagnóstico certero: en el primero de ellos se monitorea al paciente durante 48 hs mientras duerme, y bajo condiciones de fotoestimulación e hiperventilación (circunstancias en las que se incrementa la probabilidad de ocurrencia de ciertos tipos de epilepsia). Cuando se realiza un segundo EEG, se utilizan electrodos especiales, esfenoideales o temporales, y privación del sueño. Si no se detectan anomalías una última aproximación consiste en someter al paciente a video-monitoreo y electroencefalografía continuos durante un tiempo prolongado. Otro factor importante en el diagnóstico es el **examen de la historia familiar del paciente** en busca de antecedentes de epilepsia en familiares directos [3].



Sin embargo, las manifestaciones de algunos tipos menos comunes y menos diagnosticados de epilepsia pueden ser sutiles, y por lo tanto más difíciles de diferenciar de otras condiciones con sintomatología similar, y distan mucho de los síntomas de epilepsia instalados en el imaginario popular. Como ejemplos podemos mencionar las epilepsias nocturnas del lóbulo frontal (ADNFLE y epilepsia nocturna esporádica del lóbulo frontal -NFLE-), un tipo raro de epilepsia infantil con foco en el lóbulo frontal cuyas manifestaciones ocurren durante el sueño (cuando los padres o familiares del paciente muy bien pueden no estar observando al niño) [10,11]. Las mismas suelen consistir en movimientos complejos, despertares paroxísticos, gritos y hasta llanto, deambulares nocturnos e incluso incontinencia. Esto hace que sean confundidos habitualmente con pesadillas o sonambulismo, de los cuales se diferencian fundamentalmente en su periodicidad, las características de los movimientos y la presencia de actividad eléctrica anormal a nivel del SNC. La disrupción constante de la arquitectura del sueño hace que el rendimiento físico e intelectual del niño sea menor al normal, por lo que suelen ser los maestros los primeros en detectar la presencia de una patología.

Algunos casos de epilepsia parcial originada en el lóbulo temporal mesial tienen por único síntoma o síntoma predominante el miedo ictal (que puede ir desde trazas de ansiedad hasta terror intenso durante un período de treinta a sesenta segundos), diagnosticándose habitualmente un origen psiquiátrico en lugar de neurológico [12].

Dificultades en el Tratamiento

En la actualidad se reconocen varias alternativas para el tratamiento de la epilepsia [13]:



1. Farmacoterapia
2. Intervención quirúrgica
3. Estimulación del nervio vago
4. Dieta cetogénica

La **farmacoterapia** constituye el tratamiento de primera elección. Alrededor del 70% de los nuevos casos de epilepsia, tanto en niños como en adultos, pueden controlarse adecuadamente con los fármacos antiepilépticos conocidos. Los mecanismos a través de los cuales las drogas antiepilépticas protegen de las crisis son varios y se presentan esquemáticamente en la fig. 2.1 [14-19]: (1) modulación de canales iónicos voltaje-operados (canales de sodio y canales de calcio); (2) potenciación de la acción inhibitoria del ácido γ -aminobutírico (GABA), incluyendo un incremento de la inhibición sináptica mediada por los receptores GABA_A, inhibición de la recaptación de GABA desde la neurona presináptica e inhibición irreversible de la enzima GABA-transaminasa responsable del catabolismo del GABA e; (3) inhibición de la excitación mediada por receptores ionotrópicos de glutamato.

Luego de entre 2 y 5 años de respuesta exitosa al tratamiento las drogas pueden retirarse sin reaparición de las crisis epilépticas en un 70% de los niños y un 60% de los adultos [4]. Sin embargo, podemos mencionar varias problemáticas relacionadas con la farmacoterapia de la epilepsia [14]:

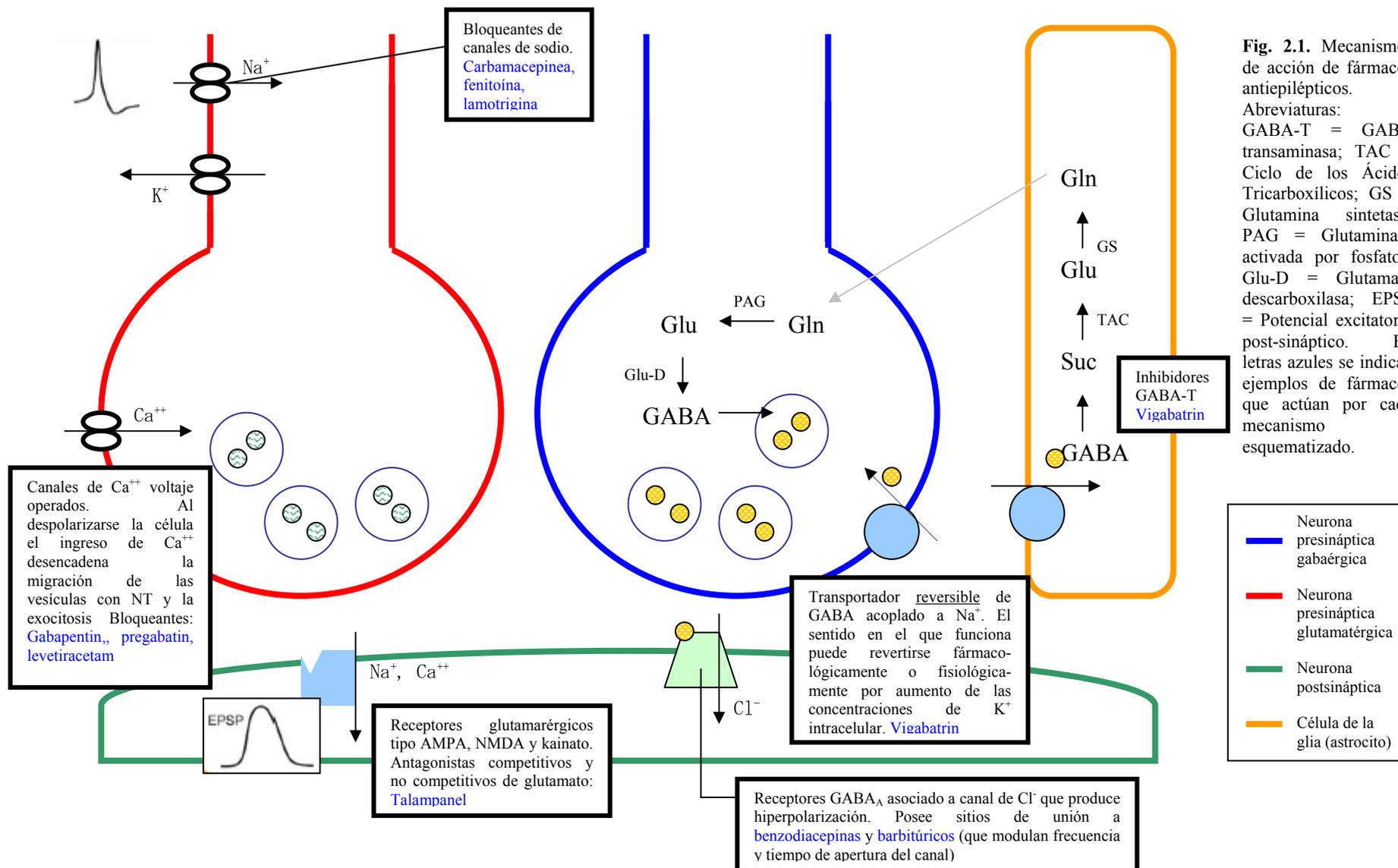
- Por razones sociales y económicas, tres cuartos de la población epiléptica mundial no tiene acceso a la medicación adecuada
- Un 30% de los pacientes epilépticos sufren de **epilepsia refractaria** o **intratable**, como se denomina a aquellos



pacientes que no responden a los agentes terapéuticos disponibles. La epilepsia refractaria es más frecuente en niños con un síndrome epiléptico criptogénico, en aquellos con un número grande de crisis antes de comenzar el tratamiento y en los que no responden al primer régimen terapéutico. Ha sido explicada desde el punto de vista biológico mediante dos hipótesis: la primera de ellas apunta a una **resistencia farmacodinámica** vinculada al blanco molecular de la droga; la segunda, a una **resistencia de tipo farmacocinética** similar a la que se puede observar en los tratamientos oncológicos. Aunque se reconoce que la resistencia a drogas puede ser multifactorial², las investigaciones más recientes se enfocan en el papel de la glicoproteína P, una bomba de eflujo de drogas ATP-dependiente, en la resistencia a los fármacos antiepilépticos. La glicoproteína P está presente en tejidos normales, particularmente en aquellos responsables de la protección del organismo frente a xenobióticos tóxicos (por ejemplo, en las células intestinales y en la barrera hematoencefálica); sin embargo, la expresión de la misma podría aparentemente incrementarse tras la exposición prolongada a farmacoterapia;

- **Efectos adversos del arsenal terapéutico disponible.** Aún la nueva generación de fármacos antiepilépticos posee efectos adversos graves tales como: sedación, ataxia, diplopía, discrasias sanguíneas graves (anemia aplástica y hemolítica, trombocitopenia), náusea, vómitos, alergia, cambios de humor, psicosis, alopecia, aumento de peso. Estas reacciones a la medicación limitan seriamente su uso, son causa frecuente de abandono del tratamiento y condicionan severamente la calidad

² Se pueden citar, entre los múltiples factores que pueden contribuir a la farmacorresistencia: modificaciones en el blanco molecular y en sistemas de recaptación, así como incrementos del metabolismo y en los mecanismos de eflujo de la droga.





de vida del paciente, aún en los casos en los que se logra controlar efectivamente la sintomatología [20].

- **Falta de fármacos antiepilépticos que prevengan la aparición o la progresión de la enfermedad.** Como ya se ha explicado, con excepción de los casos de origen genético, la epilepsia surge como respuesta a un insulto al cerebro. Si ese insulto inicial es severo, puede conducir a muerte celular isquémica; cuando es menos severo, a muerte celular apoptótica, reorganización de circuitos neuronales, remodelado de la matriz extracelular y las membranas celulares de las células del SNC o en cambios en la expresión o función de canales iónicos voltaje-operados (canales de Na^+ , Ca^{++}) o de receptores (de glutamato ionotrópicos, GABA_A). En este último caso aparece un período de latencia entre el insulto y la manifestación de la enfermedad, en el que se desarrolla un proceso conocido como **epileptogénesis**. Actualmente, Si bien se investigan los mecanismos moleculares de la epileptogénesis y el potencial antiepileptogénico de las drogas antiepilépticas conocidas y de agentes neuroprotectores (utilizando modelos crónicos de epilepsia), no existen todavía terapias basadas en la administración de agentes antiepileptogénicos luego del evento gatillo con el objeto de prevenir el desarrollo de epilepsia en grupos de riesgo (por ej., pacientes que han sido sometidos a cirugía del encéfalo, personas que ha sufrido un derrame cerebral, neonatos que han padecido encefalopatía hipóxica). Por otro lado, la epilepsia está asociada a deterioro cognitivo y cambios negativos en el comportamiento¹; en algunos casos, la frecuencia e intensidad de las crisis epilépticas aumenta a

¹ Los cambios en el comportamiento también pueden deberse, al menos parcialmente, a la ansiedad anticipatoria de ese evento súbito e imprevisible que es la crisis epiléptica (recordemos que el paciente vive con la incertidumbre de cuándo ocurrirá un nuevo episodio) y una sensación de impotencia que puede devenir en depresión mayor. Es decir, es difícil establecer cuánto del cambio comportamental debe atribuirse a motivos neurológicos y cuánto a motivos psicológicos.



medida que pasa el tiempo. Este deterioro, esta **progresión** de la enfermedad, puede ser consecuencia del desorden neurológico subyacente que causó la patología, de crisis epilépticas frecuentes y prolongadas, de anormalidades en el EEG (la interrupción reiterada del ritmo intrínseco del cerebro puede afectar las funciones de aprendizaje y memoria) o de la terapia usada en su tratamiento (ver el próximo ítem). Existe evidencia de que los trastornos del humor son no sólo un factor de riesgo de crisis epilépticas, sino uno de los factores que podría favorecer la respuesta inadecuada a la medicación. Se ha sugerido que es conveniente tener en cuenta el perfil psicotrópico de las drogas anticonvulsivas en el momento de elección del tratamiento (algunas drogas como el ácido valproico poseen efectos estabilizadores del ánimo) o complementar el tratamiento de la epilepsia con drogas antidepresivas o ansiolíticas. ¿Debería incluirse la capacidad de estabilizar el ánimo del paciente como nuevo requisito de la droga anticonvulsiva ideal?

La **intervención quirúrgica** se aplica sólo a aquellos pacientes con focos epileptiformes favorables a la resección. Si bien erradica las crisis epilépticas en epilepsias parciales farmacorresistentes en alrededor del 50% de los pacientes operados (este porcentaje varía según en qué zona del cerebro se halle el foco; la eficacia a largo plazo es del 70% en el caso de resecciones del lóbulo temporal y de un 30% en el caso del lóbulo frontal) [21], presenta morbosidad alta, sobre todo en pacientes adultos, que incluye afasia, hemiparesia, hemianopia, déficit motor postoperatorio, deterioro de las capacidades cognitivas (capacidad de atención, memoria y aprendizaje), hemorragias recurrentes del SNC e infecciones. Es, por otro lado, un tratamiento costoso que no todos los pacientes pueden afrontar. La **estimulación**



del **nervio vago** (aprobada por la Food and Drug Administration para el tratamiento de la epilepsia, en 1997, y de la depresión, en 2005) consiste en la implantación en el pecho del paciente de un dispositivo electrónico que estimula el nervio vago con pulsos de entre 0.25 y 1.50 mA a intervalos programados por el médico. Es menos eficaz que la cirugía (logra reducir la intensidad o frecuencia de las crisis, pero no erradicarlas, y se utiliza como terapia adjuntiva, es decir, no reemplaza la medicación pero mejora sus resultados y permite reducir la dosis de la misma). A pesar de no funcionar como monoterapia, la estimulación vagal presenta efectos adversos menores: anormalidades en la voz, tos, ronquera, y apnea suave del sueño. No obstante el dispositivo es costoso y esto limita su uso, habiendo sólo unos 20,000 pacientes en el mundo que lo utilizan [22]. La **dieta cetogénica** consiste en inducir un estado de cetosis limitando la cantidad de proteínas y carbohidratos que ingiere el paciente, de manera que los lípidos se constituyan en la principal fuente de energía del organismo [23]. Se han postulado varias explicaciones para el uso de esta alternativa en el tratamiento de la epilepsia. Aparentemente, los cuerpos cetónicos poseen un efecto estabilizante sobre el SNC, y el estado de acidosis y los cambios en el balance electrolítico que aparecen como consecuencia de la dieta podrían modificar el umbral convulsivo. Si bien los estudios demuestran que la cantidad de pacientes en los que desaparecen las convulsiones es muy inferior al que encontramos en el caso de la farmacoterapia y la cirugía, la dieta cetogénica **puede utilizarse como terapia de todos los tipos de epilepsia**. Presenta también efectos adversos graves, tales como retraso en el desarrollo y en el crecimiento en niños e hipercolesterolemia e hipertrigliceridemia. La tabla 2.2 muestra una síntesis de las ventajas y desventajas de cada una de estas alternativas terapéuticas.



De lo anteriormente expuesto se desprende la necesidad urgente de nuevos tratamientos para el paciente epiléptico. En el caso particular de la farmacoterapia, se requieren drogas que presenten los siguientes requisitos ideales [14]:

- ausencia de fármaco-resistencia;
- inhibición de la epileptogénesis.
- baja toxicidad;
- mayor accesibilidad por parte de la población, es decir, bajo costo;
- efectos estabilizadores del ánimo para inhibir cambios negativos en el comportamiento del paciente propios de la progresión de la enfermedad (tanto por causas psicológicas como neurológicas);

Tabla 2.2. Ventajas y desventajas de los distintos tratamientos de la epilepsia.

Tratamiento	Ventajas	Desventajas
Farmacoterapia	<ul style="list-style-type: none">• La farmacoterapia logra controlar la sintomatología en alrededor del 70% de los pacientes. En un 60% de los adultos y un 70% de los niños cuya sintomatología es controlada adecuadamente por los fármacos se produce la remisión de la enfermedad (es decir, entre un 40 y 50% de la población epiléptica puede curarse mediante farmacoterapia).• Si bien existen efectos adversos graves, la mayoría de ellos desaparecen al retirar la medicación.	<ul style="list-style-type: none">• Un 30% de la población epiléptica es farmacorresistente.• No hay evidencia de que existan fármacos que actúen inhibiendo la epileptogénesis o inhiban la progresión de la enfermedad.• Efectos adversos graves que reducen la calidad de vida del paciente aún cuando se controle totalmente la sintomatología.• Un 70-75% de la población epiléptica no tiene acceso a la medicación.
Cirugía	<ul style="list-style-type: none">• Alto grado de efectividad: erradica las crisis en un	<ul style="list-style-type: none">• Alto costo.• Morbosidad: si bien en



	30-70% de los pacientes operados según el lugar afectado.	baja proporción, las secuelas suelen ser graves e irreversibles.
	<ul style="list-style-type: none">• Es efectivo en las epilepsias parciales, que representan la mayoría de los casos de epilepsia refractaria.	<ul style="list-style-type: none">• No todos los pacientes son elegibles para intervención quirúrgica.• No permite tratar las epilepsias generalizadas.
Estimulación del nervio vago	<ul style="list-style-type: none">• Efectos adversos suaves	<ul style="list-style-type: none">• Alto costo.• Es terapia adyuvante.• No logra la remisión de la enfermedad.
Dieta Cetogénica	<ul style="list-style-type: none">• Bajo costo.	<ul style="list-style-type: none">• Efectos adversos graves.• Disminución de la calidad de vida asociada a las importantes restricciones en la dieta.

En este trabajo de tesis nos enfocaremos en la búsqueda racional de nuevos fármacos anticonvulsivos activos en el ensayo MES mediante aplicación de descriptores de baja dimensionalidad (constitucionales y topológicos, ver Caps. 2 y 3) en la selección de compuestos promisorios desde grandes bases de datos/colecciones virtuales de estructuras químicas. Las aplicaciones de descriptores topológicos en SV permiten detectar estructuras novedosas, ya que el descriptor topológico, como se verá, puede capturar características sutiles de la estructura molecular, **permitiendo seleccionar estructuras que retengan las características esenciales que determinan la actividad biológica buscada pero en un contexto estructural diferente, novedoso.** En otras palabras, se parte de lo ya conocido para encontrar actividad en entidades estructurales no conocidas. Por otro lado, el SV permite encontrar compuestos activos abundantes en la naturaleza o de síntesis o hemisíntesis sencilla, hecho que determina eventualmente una reducción de costos de producción del medicamento antiepiléptico. Esta posibilidad brindará una mayor accesibilidad de la población al tratamiento, que como se ha visto es fundamental en los países



subdesarrollados. Debe mencionarse además que ofrece una mayor probabilidad de éxito en ensayos clínicos y preclínicos, teniendo en cuenta para ello el uso de filtros ADME/Tox en la selección de estructuras mediante SV. Este recurso contribuye a optimizar los recursos invertidos en el desarrollo de fármacos, lo cual también favorece el desarrollo de agentes terapéuticos a menor costo.

Se verá que las nuevas drogas descubiertas mediante la metodología propuesta son, según el caso y respondiendo a estos objetivos: 1) estructuralmente diversas y con características estructurales no observadas en el arsenal antiepiléptico ya conocido; 2) productos naturales abundantes en la naturaleza; 3) estructuras sencillas de fácil síntesis.

El modelo animal utilizado en los ensayos biológicos es un modelo agudo (inducción de la convulsión en ratones mediante un estímulo cuya magnitud está por encima del umbral convulsivo del animal), por lo tanto la tesis no abordará el problema de la epileptogénesis pese a que el grupo de trabajo ha iniciado ya investigaciones referidas a esta problemática mediante la metodología aplicada en el presente trabajo [24].

Referencias.

1. World Health Organization, fact sheet nro. 265. Mental and neurological disorders. **2001**
2. Fisher, R.S.; van Emde Boas, W.; Blume, W.; Elger, C.; Genton, P.; Lee, P.; Engel Jr., J. Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia*. **2005**, 46, págs. 470-472
3. Browne, T.R.; Holmes, G.L. Epilepsy. *N. Engl. J. Med.* **2001**, 344, págs. 1145-1151



4. World Health Organization, fact sheet nro. 165. Epilepsy: aetiology, epidemiology and prognosis. **2001**
5. World Health Organization, fact sheet nro. 166. Epilepsy: social consequences and economic aspects. **2001**
6. Commission on Epidemiology and Prognosis, International League Against Epilepsy. Guidelines for Epidemiologic Studies on Epilepsy. *Epilepsia*. **1993**, 34, págs. 592-596
7. Mulley, J.C.; Scheffer, I.E.; Petrou, S.; Berkovic, S.F. Channelopathies as a genetic cause of epilepsy. *Curr. Opin. Neurol.* **2003**, 16, págs. 171-176
8. Aridon, P.; Marini, C.; Di Resta, C.; Brilli, E.; De Fusco, M.; Politi, F.; Parrini, E.; Manfredi, I.; Pisano, T.; Pruna, D.; Curia, G.; Cianchetti, C.; Pasqualetti, M.; Becchetti, A.; Guerrini, R.; Casari, G. Increased sensitivity of the neuronal nicotinic receptor $\alpha 2$ subunit causes familial epilepsy with nocturnal wandering and ictal fear. *Am. J. Hum. Genet.* **2006**, 79, págs. 342-350
9. Buelow, J.M.; Shore, C.P. The treatment of chronic epilepsy: a review of recent studies of clinical efficacy and side effects. *Epilepsy Behav.* **2006**, 9, págs. 440-447
10. Combi, R.; Dalprà, L.; Tenchini, M.L.; Ferini-Strambi, L. Autosomal dominant nocturnal frontal lobe epilepsy. A critical overview. *J. Neurol.* **2004**, 251, págs. 923-934
11. Zucconi, M.; Ferini-Strambi, L. NREM parasomnias: arousal disorders and differentiation from nocturnal frontal lobe epilepsy. *Clin. Neurophysiol.* **2000**, 111 (Supl. 2), págs. S129-S135
12. Pellegrino Rosa, V.; de Araújo Filho, G.M.; Andriani Rahal, M.; Sales Ferreira Caboclo, L.O.; Ceiki Sakamoto, A.; Yacubian, E.M.T. Ictal Fear: Semiologic Characteristics and Differential Diagnosis with Interictal Anxiety Disorders. *J. Epilepsy Clin. Neurophysiol.* **2006**, 12, págs. 89-94
13. Shorvon, S. The treatment of chronic epilepsy: a review of recent studies of clinical efficacy and side effects. *Curr. Opin. Neurol.* **2007**, 20, 159-163
14. Löscher, W.; Schmidt, D. New horizons in the development of antiepileptic drugs: innovative strategies. *Epilepsy Res.* **2006**, 69, págs. 183-272
15. Siegel, G.J. (editor). Basic neurochemistry. Molecular, cellular and medical aspects. Lippincott-Raven Publishers, Philadelphia, USA, 6ta. edición, **1999**, págs. 335-345
16. Bak, L.K.; Schousboe, A.; Waagepetersen, H.S. The glutamate/GABA-glutamine cycle: aspects of transport, neurotransmitter homeostasis and ammonia transfer. *J. Neurochemistry.* **2006**, 98, págs. 641-653
17. Wu, Y.; Wang, W.; Richerson, G.B. Vigabatrin induces tonic inhibition via GABA transporter reversal without increasing vesicular GABA release. *J. Neurophysiol.* **2003**, 89, págs. 2021-2034



18. Schousboes, A.; Sarup, A.; Larsson, O.M.; White, H.S. GABA transporters as drug targets for modulation of GABAergic activity. *Biochem. Pharmacol.* **2004**, 68, págs. 1557-1563
19. Wu, Y.; Wang, W.; Richerson, G.B. GABA Transaminase Inhibition Induces Spontaneous and Enhances Depolarization-Evoked GABA Efflux via Reversal of the GABA Transporter. *J. Neurosci.* **2001**, 21, págs. 2630-2639
20. Bialer, M.; Johannessen, S.I.; Kupferberg, H.J.; Levy, R.H.; Perucca, E.; Tomson, T. Progress report on new antiepileptic drugs: a summary of the seventh Eilat Conference (EILAT VII). *Epilepsy Res.* **2004**, 61, págs. 1-48
21. Tellez-Centeno, J.F.; Dhar, R.; Wiebe, S. Long-term seizure outcomes following epilepsy surgery: a systematic review and meta-analysis. *Brain.* **2005**, 128, págs. 1188-1198
22. Schachter, S.C. Therapeutic effects of vagus nerve stimulation in epilepsy and implications for sudden unexpected death in epilepsy. *Clin. Auton. Res.* **2006**, 16, págs. 29-32
23. Lefevre, F.; Aronson, N. Ketogenic diet for the treatment of refractory epilepsy in children: a systematic review of efficacy. *Pediatrics*, **2000**, 105, págs. 46-52
24. Bellera, C.L.; Talevi, A.; Bruno-Blanch, L.E. Aplicación de análisis lineal discriminante en la búsqueda de drogas antiepileptogénicas. *Latin American J. Pharm.* **2007**, 26, págs. 244-252



Capítulo 2. Desarrollo de nuevos fármacos.

- Métodos clásicos y modernos.
- Relaciones Cuantitativas Estructura-Propiedad y Estructura-Actividad.
- Clasificación de aproximaciones QSAR.
- Fundamentos de la metodología elegida para la presente tesis.
Etapas generales de un Estudio QSPR/QSAR.

Métodos clásicos y modernos.

El objetivo central de este capítulo es realizar una descripción de las metodologías QSAR. Se describirán en primera instancia las distintas metodologías empleadas en el pasado y en el presente para el desarrollo de nuevos fármacos. Las distintas estrategias pueden clasificarse en dos grandes grupos: **métodos clásicos** y **métodos modernos**.

Dentro de los métodos clásicos pueden incluirse el **screening sistemático** y la **modificación estructural de compuestos líderes** (algunos autores mencionan dentro de los métodos clásicos a la serendipicidad, pero como se vio difícilmente pueda ser considerada como una verdadera estrategia o método).

El **screening sistemático**, como ya se ha mencionado, se popularizó luego de que Fleming descubriera la penicilina. El escenario histórico impuesto por la Segunda Guerra Mundial favorecía el interés de los gobiernos de las grandes potencias por encontrar nuevos antibióticos, pero casi no existían conocimientos científicos vinculados a esta temática. Sin antibióticos, muchos soldados morían en el frente como consecuencia de infecciones de heridas, incluso



heridas relativamente leves (hasta entonces la primera causa de muerte de heridos era la mionecrosis clostridiana) por lo que el descubrimiento de nuevos antibióticos implicaba una clara ventaja estratégica. Fueron Waksman y su grupo de colaboradores quienes, buscando antibióticos capaces de inhibir el crecimiento de bacterias Gram negativas (la penicilina sólo se había demostrado eficaz contra bacterias Gram positivas) encararon la búsqueda de antibióticos de una manera nueva, sistemática, buscando zonas de inhibición alrededor de colonias de una serie de microorganismos del suelo, en un principio, y de bacterias patogénicas luego. Miles de bacterias diferentes fueron aisladas y ensayadas en busca de actividad antibiótica. El éxito de esta metodología de prueba y error (Waksman descubrió unos veinte nuevos agentes antimicrobianos naturales, entre ellos la actinomicina y la estreptomycinina) hizo que prontamente el screening sistemático se realizara sobre otras fuentes de productos naturales en busca no sólo de antibióticos sino de fármacos pertenecientes a otras categorías terapéuticas [1]. Dado que a través de la Medicina Tradicional las drogas vegetales habían demostrado ser una cuantiosa fuente de principios activos¹, los productos naturales se convirtieron en un nuevo blanco del screening sistemático. Este screening sistemático prontamente se asoció al fraccionamiento bioquímico, en el cual la droga vegetal es extraída sucesivamente con solventes de polaridad creciente, y sobre cada extracto se realizan ensayos biológicos para determinar el extracto activo a partir del cual seguidamente se procederá al aislamiento del o los principios activos. Recientemente el screening sistemático ha sido aplicado extensiva y exitosamente a productos naturales de origen marino, tales como algas, bacterias o esponjas marinas [2-4].

¹ Como afirma Graham Patrick en su Introducción a la Química Medicinal (Oxford University Press, 2001) “la evolución ya ha llevado a cabo su propio proceso de screening, por el cual las plantas tienen más chances de sobrevivir si contienen potentes compuestos que evitan que animales e insectos las coman”.



La **modificación estructural** constituye la otra estrategia clásica para el desarrollo de nuevos fármacos, orientada a obtener, a partir de compuestos activos, otros con otras características tales como mayor potencia, mayor índice terapéutico (LD_{50}/ED_{50}), mayor biodisponibilidad o, simplemente, encontrar un nuevo compuesto activo no patentado. El origen del prototipo activo sobre el que se realiza la modificación puede ser muy variado: un ligando natural, una droga patentada por una compañía farmacéutica competidora (drogas “me too” / “semejante a”), un compuesto hallado por serendipidad, screening sistemático, observación de un efecto adverso conveniente, etc [5]. Dentro de esta estrategia encontramos a su vez distintos tipos de modificaciones clásicas tales como **homologación**, **ramificación**, **ciclación**, **apertura de anillos**, **adición de un escudo estérico**, **transformaciones que involucran anillos** y **modificaciones guiadas por el concepto de bioisosterismo**.

Por **homologación** entendemos la generación de una serie de derivados de un compuesto por introducción de unidades constantes (como por ejemplo, grupos metileno). Hasta un valor crítico de unidades adicionadas (que varía según la actividad considerada) se observa un aumento de la actividad por efectos farmacocinéticos y farmacodinámicos. El aumento progresivo de grupos metilo incrementa la lipofilia, lo cual favorece la permeabilidad a través de membranas biológicas, no obstante más allá de un valor crítico la actividad comienza a decrecer por ser el compuesto incapaz de disolverse o por quedar retenido el mismo en las barreras biológicas (sin distribuirse); el efecto farmacodinámico se relaciona con la presencia de bolsillos hidrofóbicos en el receptor biológico: cuando la parte hidrofóbica excede el tamaño del bolsillo se desfavorece la interacción-droga receptor. Otra explicación de este efecto es que



usualmente el ligando debe perder el agua de solvatación para interactuar con el sitio activo, proceso que se favorece en moléculas con cierta hidrofobicidad.

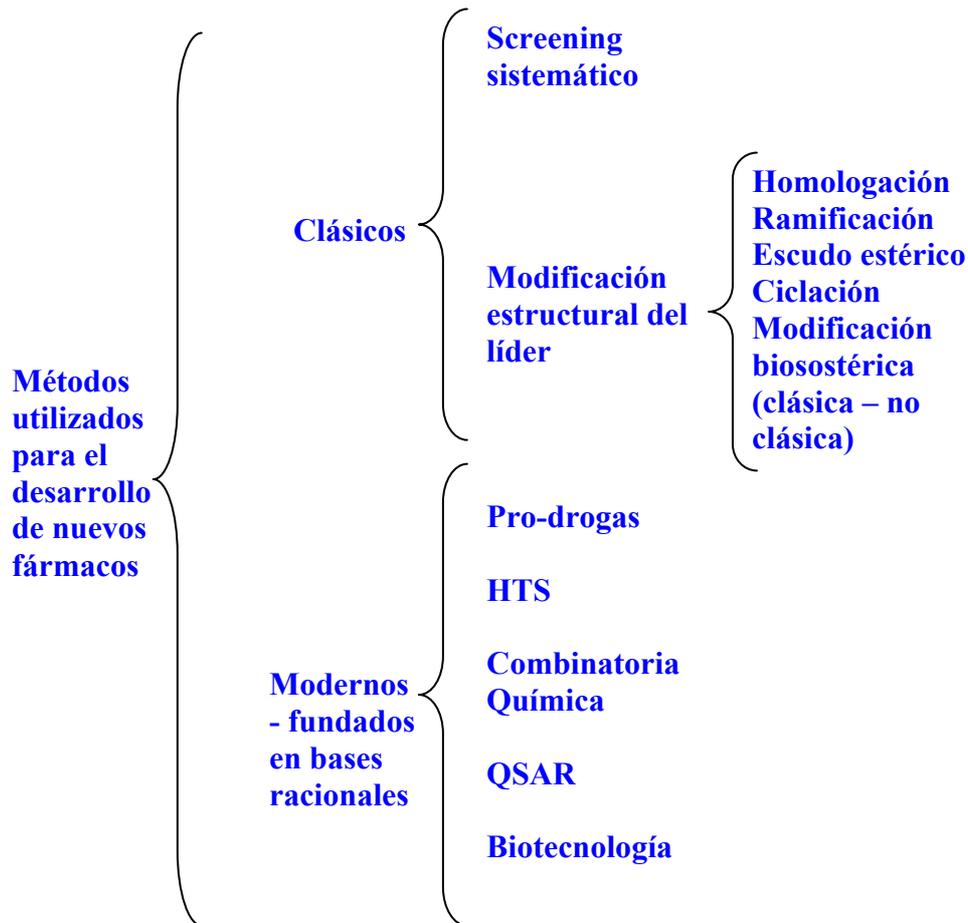


Fig. 3.1 Distintos métodos utilizados en el desarrollo de fármacos.

La **ramificación** usualmente disminuye o cambia la actividad por desfavorecer la interacción con un receptor determinado. Un ejemplo clásico son la prometazina y la promazina (fig. 3.2) [6]. La primera tiene, predominantemente, acción atihistamínica y antiespasmódica, mientras que la segunda posee acción antipsicótica.



La **adición de un escudo o impedimento estérico** se refiere a la protección de funciones químicas susceptibles de degradación química o enzimática (por ejemplo, grupos éster o grupos amida) introduciendo grupos voluminosos (por ejemplo, un sustituyente terbutilo) para “esconder” el grupo lábil del ataque del nucleófilo o de la enzima [5].

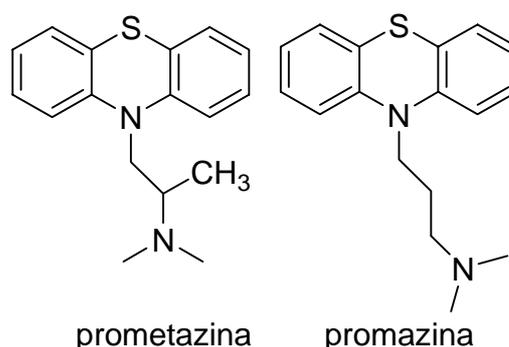


Fig. 3.2. Ejemplo de modificación del perfil farmacológico de un fármaco mediante ramificación.

Dentro de las **transformaciones que involucran anillos** encontramos la apertura o cierre de anillos, la contracción o expansión de anillos (cambio de un anillo por otro con menos o más átomos, respectivamente) y la reorganización de anillos, entre varias otras. Por ejemplo, la **ciclación o cierre de anillos** supone la transformación de un sustituyente alquílico en un análogo cíclico. Aunque puede introducir pequeños cambios en la lipofilia o en el metabolismo de la molécula, en general no produce grandes cambios en la actividad. Por ejemplo, la clorpromazina y el análogo NSC17471 obtenido mediante ciclación poseen acción antipsicótica similar (Fig. 3.3) mientras que por su parte la trimeprazina y la metdilazina poseen actividades antihistamínicas parecidas.

Una estrategia ampliamente utilizada para la modificación estructural de un compuesto líder se basa en el concepto de



biososterismo [7]. En 1919 Langmuir, estudiando el comportamiento químico y la reactividad de un grupo de compuestos con igual número de electrones de valencia acuñó el concepto de **isosterismo** para definir átomos y/o moléculas inorgánicas u orgánicas que poseen el mismo número y/o arreglo de electrones (por ejemplo, CO_2 y NO_2 , PO_4^{-3} y SO_4^{-2}).

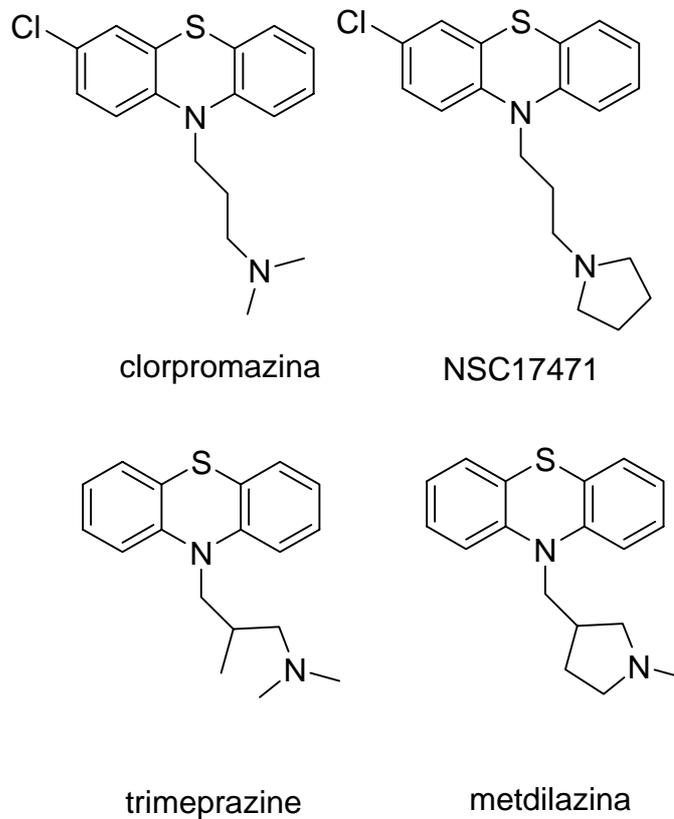


Fig. 3.3. La ciclación suele modificar el perfil farmacocinético de un fármaco sin afectar su perfil farmacológico.

En 1925 Grimm formuló la **Ley del Desplazamiento**, una regla empírica que establece que la adición de entre uno y cuatro átomos de hidrógeno a un átomo de los cuatro grupos anteriores a los gases inertes en la tabla periódica da como resultado un pseudoátomo de propiedades similares a las de los cuatro grupos a su derecha. La Ley del Desplazamiento se ilustra en la tabla 3.1.



Tabla 3.1. Ley del Desplazamiento de Grimm. Los átomos y pseudoátomos de una misma columna se comportan como isósteros.

C	N	O	F	Ne	Na
	CH-	NH-	OH-	FH	-
		CH ₂ -	NH ₂ -	H ₂ O	FH ₂ ⁺
			CH ₃ -	NH ₃	H ₃ O ⁺
				CH ₄	NH ₄ ⁺

Más adelante, Erlenmeyer propuso una definición más amplia de isosterismo, definiendo isósteros como elementos, moléculas o iones que presentan el mismo número de electrones en la capa de valencia. Tiempo después Friedman reconocería la utilidad del concepto de isosterismo para diseñar moléculas bioactivas, definiendo como bioisósteros a aquellas moléculas que, cumpliendo con la definición de isosterismo, actúan sobre el mismo receptor ya sea como agonistas o como antagonistas. En 1979 Thornber propuso una definición más amplia de bioisosterismo; según él, los bioisósteros son moléculas o grupos con propiedades fisicoquímicas similares y una actividad biológica aproximadamente parecida [8]. En 1991 Burguer propuso una nueva definición, sugiriendo que son bioisósteros aquellas moléculas o grupos que poseen un volumen y una distribución electrónica similar y propiedades fisicoquímicas parecidas [9]. El bioisosterismo clásico incluye átomos o grupos mono, di y trivalentes, átomos tetrasustituídos y anillos equivalentes; el bioisosterismo no clásico incluye además sustituciones de ciclos por grupos acíclicos y sustitución de grupos funcionales (por ejemplo, un grupo éster por un grupo amida). La fig. 3.4 presenta un ejemplo de bioisosterismo, en este caso bioisosterismo clásico entre los grupos -OH y -NH₂.

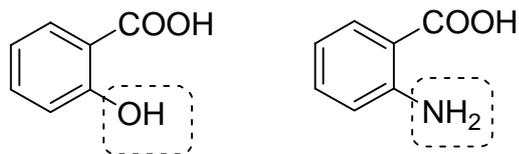


Fig. 3.4. Reemplazo bioisostérico clásico del grupo hidroxilo del ácido salicílico por el grupo amino del ácido antranílico.

Por su parte, dentro de las metodologías más modernas encontramos el uso de pro-drogas, la combinatoria química, el screening de alta performance o HTS, las metodologías QSAR y las aplicaciones de la biotecnología en el área de la Química Medicinal.

Por **pro-droga** entendemos un tipo de compuesto que debe ser metabolizado en el organismo para poder ejercer su acción biológica. En general, las pro-drogas consisten de un fármaco unido a un grupo modulador, pudiendo el grupo modulador ser escindido mediante alguno de los sistemas metabólicos del organismo y liberándose el fármaco para poder ejercer su acción terapéutica (siendo que la asociación fármaco-grupo modulador es, en el caso del pro-fármaco ideal, farmacológicamente inactiva) [10]. Las pro-drogas pueden ser diseñadas para resolver problemas farmacocinéticos (por ejemplo, baja absorción) o dificultades debidas a la toxicidad, estabilidad o baja aceptación por parte del paciente (como ser, características organolépticas desfavorables). La levodopa constituye un ejemplo de pro-fármaco que permite superar la baja permeabilidad de la dopamina a través de la BBB. La levodopa es transportada activamente a través de la barrera y luego es metabolizada a dopamina por una descarboxilasa del SNC, ejerciendo su acción terapéutica (fig. 3.5)

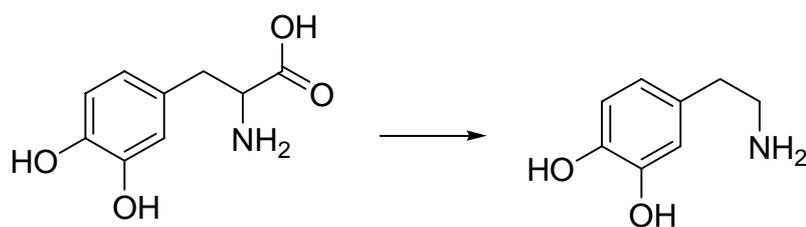


Fig. 3.5. La levodopa, profármaco de la dopamina, logra atravesar la BBB mediante transporte activo.

El **screening de alta performance** (HTS) combina robótica y microensayos para ensayar biológicamente un gran número de compuestos con el fin de verificar su actividad contra un gran número de blancos moleculares. Típicamente, miles de compuestos pueden ser ensayados simultáneamente en medio centenar de ensayos a la vez. Se trata de combinar el screening sistemático clásico de Waksman con la tecnología moderna ensayando en poco tiempo, y con pequeña cantidad de muestra, miles de compuestos en una gran cantidad de blancos moleculares potenciales. La aparición del HTS y los avances en el mapeo del genoma humano (que trajo aparejado el descubrimiento de un gran número de blancos moleculares inexplorados) hicieron necesario desarrollar nuevas tecnologías de síntesis que permitieran sintetizar una gran diversidad de nuevos compuestos en poco tiempo. Surgió así la **síntesis combinatoria**, que consiste en llevar a cabo reacciones definidas (típicamente en fase sólida, aunque se están desarrollando tecnologías de síntesis combinatoria en fase líquida) a partir de reactivos de partida diversos en un solo reactor (síntesis compleja) o en reactores separados (síntesis en paralelo) [11,12].

Las metodologías **QSAR** se describirán en detalle en los apartados subsiguientes del presente capítulo. Las **aplicaciones de la biotecnología en la Química Medicinal** incluyen la llamada terapia génica (introducir un gen funcional en las células de un paciente con falta o deficiencia de un gen) y la farmacogenómica (que consiste en



utilizar información genética para predecir la eficacia y seguridad de una droga en un paciente o grupos de pacientes determinado). Adicionalmente, la biotecnología ha aportado a la Química Medicinal animales transgénicos que pueden utilizarse como modelos crónicos y ha identificado y elucidado la estructura de nuevos blancos moleculares, estableciendo nuevos objetivos terapéuticos. Permite además aportar información sobre los mecanismos moleculares de patologías de origen genético y facilita el diseño de nuevas estructuras mediante metodologías basadas en las interacciones específicas fármaco-receptor (docking), generando nuevos fármacos con características complementarias a las del sitio activo del receptor.

Relaciones Cuantitativas Estructura-Propiedad y Estructura-Actividad.

El proceso de desarrollo de un fármaco consistía tradicionalmente en identificar una estructura con un perfil farmacológico interesante (prototipo activo), sobre la que se llevaban a cabo una **serie de modificaciones estructurales con el objeto de optimizar la actividad**², mientras paralelamente se investigaban **rutas y condiciones de síntesis alternativas**, con objetivos diversos tales como mejorar los rendimientos, disminuir los costos, utilizar reactivos y solventes de menor toxicidad o facilitar los procedimientos de control de reacción y aislamiento y purificación de los productos obtenidos.

Una vez que una serie de estructuras relacionadas (variación estructural exhaustiva del prototipo activo) habían sido sintetizadas y ensayadas farmacológicamente, el químico medicinal estaba en condiciones de esgrimir hipótesis que relacionaran las características estructurales y las propiedades fisicoquímicas de los compuestos que

² En primer término, se proponían una serie de modificaciones sintéticamente accesibles según la estructura del compuesto de partida.



conformaban la serie con sus correspondientes actividades biológicas. A partir de estas hipótesis era posible encarar una segunda serie de optimizaciones estructurales con fundamentos racionales. Por lo tanto, las primeras relaciones estructura-actividad (REA) se basaban en la observación visual de representaciones sencillas de la estructura molecular por parte de un químico experto. La serie de compuestos inspeccionados estaba constituida generalmente por compuestos íntimamente relacionados desde el punto de vista estructural (ya que surgían de pequeñas modificaciones de la estructura química del fármaco que constituía el punto de partida). Se desprende que las primeras REA estaban limitadas por:

- a) basarse en representaciones simples de la estructura molecular (estructuras en el plano o modelos moleculares sencillos); no se contaba con representaciones estructurales válidas;
- b) establecerse a partir de inspección visual de las estructuras comprendidas por la serie, hecho que dificultaba encontrar relaciones no evidentes entre los compuestos estudiados;
- c) estar la serie conformada por estructuras con pequeñas diferencias estructurales. En otras palabras, se trataba de relaciones parciales, que se establecían por inspección de un espacio químico pequeño, limitado; por lo tanto, las conclusiones obtenidas a partir de esa inspección podían aplicarse en un espacio químico o dominio de aplicabilidad pequeño y;
- d) conocimiento limitado de los procesos bioquímicos que definen la actividad de un fármaco.

El hallazgo fundamental de Hammett (descrito brevemente en la Introducción) consistió justamente en superar las limitaciones de la inspección visual por parte de un experto, recurriendo a métodos



estadísticos para relacionar características de la estructura molecular con una propiedad fisicoquímica de una serie de compuestos:

$$\log \frac{K}{K_o} = \rho \log \frac{K'}{K'_o}$$

donde como ya se describió, K_o y K'_o representan, respectivamente, las constantes de disociación del ácido fenilacético y el ácido benzoico sin sustituir, y K y K' las constantes de disociación de los mismos ácidos, en ese orden, sustituidos por el mismo sustituyente en la misma posición del anillo aromático. La expresión anterior puede expresarse como:

$$\log \frac{K}{K_o} = \rho \sigma$$

donde σ denota la contribución electrónica del sustituyente respecto al hidrógeno. A partir de la ecuación anterior, determinando σ para diversos sustituyentes, puede estimarse el efecto que cada sustituyente tendrá sobre la función principal, sin necesidad de la determinación experimental de la misma.

Por su parte Hansch y Fujita tuvieron la lucidez y la audacia de proponer la aplicación de la aproximación de Hammett en la predicción de actividades biológicas de compuestos químicos, introduciendo el término “relaciones cuantitativas estructura-actividad” (Quantitative Structure-Activity Relationship - QSAR). Hansch y Fujita advirtieron prontamente que **al considerar la actividad biológica nos referimos a una propiedad que, aunque dependiente de la estructura química como todas las propiedades moleculares, es particularmente compleja**, por cuanto la capacidad de un compuesto químico para manifestar una



actividad biológica determinada depende a su vez de una serie de procesos complejos³ además de la interacción con su o sus blancos moleculares (receptores, enzimas), entre ellos: la absorción a través de las diferentes barreras biológicas presentes en un organismo, la metabolización a especies activas o inactivas, el reparto (no uniforme) en los distintos tejidos que componen los órganos de un sistema vivo y la velocidad de eliminación del compuesto administrado y sus metabolitos, entre otros factores. Hansch y Fujita enfrentaron los desafíos particulares que supone el modelado de la actividad biológica al intentar utilizar el parámetro σ de Hammet para predecir bioactividades, sin buenos resultados. La descripción de las características electrónicas de una serie de compuestos era, por tanto, insuficiente para explicar consistentemente las diferencias en sus actividades biológicas observadas; fue necesario la consideración de otros aspectos estructurales, tales como la lipofilia (introduciendo el parámetro π de Hansch y Fujita⁴) y los efectos estéricos (a través, por ejemplo, de la refractividad molar) [13,14] para describir propiedades biológicas de una serie de estructuras. De esta manera, la actividad biológica podía ser expresada, recurriendo a los métodos de ajuste más sencillos (regresión lineal múltiple) como:

$$\log \frac{1}{C} = a_1 \sigma + a_2 \pi + a_3 MR + a_0 \quad (2.1)$$

donde σ , π y MR son los valores del parámetro de Hammet, el parámetro de Hansch y Fujita y la Refractividad Molar para el sustituyente considerado y a_0 es el valor predicho por la ecuación para el compuesto

³ Hansch y sus colaboradores trabajaron originalmente en la predicción de la actividad biológica de auxinas (hormonas involucradas en la regulación del crecimiento en vegetales), y luego ampliaron el uso de las metodologías desarrolladas a bioactividades diversas.

⁴ Definido en forma análoga al σ de Hammet pero caracterizando la influencia de diversos sustituyentes sobre la el logaritmo de la constante de reparto octanol-agua: $\pi R = \log PR - \log PH$, donde PR y PH son, respectivamente las constantes de reparto octanol-agua del compuesto sustituido por el sustituyente R y sin sustituir



sin sustituir. C es la variable que cuantifica la actividad biológica. En general, C representa alguna de las formas en las que se suele expresar la actividad biológica de un compuesto bioactivo: ED_{50} (dosis del compuesto que genera el 50% del efecto máximo o dosis que genera el efecto deseado en la mitad de la población), IC_{50} (concentración requerida para que se produzca un 50% de inhibición) o LD_{50} (dosis que provoca muerte en el 50% de la población a la que se le administra el compuesto). La ecuación puede generalizarse:

$$\log \frac{1}{C} = a_1 (\text{descriptor electrónico}) + a_2 (\text{descriptor hidrofóbico}) + a_3 (\text{descriptor estérico}) + a_0 \quad (2.2)$$

Dado que por su definición el σ de Hammett y el π de Hansch y Fujita son expresiones de cambios de energía libre, y que según la Teoría de Receptores tradicional los valores de ED_{50} o IC_{50} corresponden a la constante de disociación del complejo droga-receptor (K_d) [15], debe expresarse la actividad biológica como $\log 1/C$ o $\log C$ para que se corresponda con el cambio de energía libre asociada a la unión de la droga al receptor o a la disociación del complejo fármaco-receptor⁵

Cabe mencionar que, si bien Hansch y sus colaboradores inicialmente pensaron a las relaciones QSAR como una metodología promisorio para guiar optimizaciones de estructuras conocidas, pronto advirtieron que la extrapolación de las relaciones obtenidas hacia regiones ubicadas fuera del espacio definido por los descriptores y las actividades biológicas del conjunto de entrenamiento conducía usualmente a predicciones inexactas del valor de actividad. En otras palabras, la actividad biológica no es en general una función estrictamente lineal de un conjunto de descriptores. Un ejemplo

⁵ Originalmente, cuando el término QSAR no estaba arraigado, este tipo de relaciones estructura-actividad eran referidas como Relaciones Lineales de Energía Libre.



clásico es un conjunto de estructuras en las que la actividad aumenta sistemáticamente con el valor del $\log P$; sería un error interpretar que un compuesto con un $\log P$ muy grande tendrá una actividad muy grande, por cuanto superado un valor crítico de $\log P$ la actividad biológica en general comenzará a descender por reducirse la solubilidad y distribución del compuesto y, consecuentemente, su biodisponibilidad. Por lo tanto, la metodología QSAR permite intrapolaciones dentro del espacio definido por el conjunto de entrenamiento pero no extrapolaciones fuera del mismo. Se desprende que no puede confiarse en las predicciones de valores de actividad fuera del rango de actividades definido por el conjunto de entrenamiento, por lo que las metodologías QSAR no resultan válidas para conducir una optimización.

Aún así, estas metodologías tuvieron un notable éxito en la predicción de valores de actividad para compuestos incluidos en el espacio químico definido por el conjunto de entrenamiento (es decir, predicciones basadas en intrapolación), y pronto se propusieron un gran número de nuevos parámetros (o descriptores) que caracterizaran cuantitativamente estos y otros aspectos de la estructura molecular. Dos pasos importantes en el desarrollo de los métodos QSAR (que como se verá, serán fundamentales para introducir luego la idea de Screening Virtual) fueron a) **la incorporación de descriptores de naturaleza “holística”** (que cuantificaban aspectos electrónicos, hidrofóbicos y estéricos no ya de un sustituyente sino de toda una molécula) y b) **la introducción de descriptores derivados de modelos moleculares teóricos** en los estudios QSAR.

El primero de estos dos adelantos (la cuantificación de propiedades moleculares en lugar de utilizar el valor de una propiedad para de un fragmento de la molécula) permitió que el QSAR dejara de ser una metodología aplicable estrictamente a series de compuestos



similares, formadas por un mismo compuesto de partida y por derivados obtenidos por incorporación a la molécula de sustituyentes con valores de σ , π , etc. conocidos.

La definición de descriptores cuyo valor, para cada compuesto, podía ser calculado teóricamente (en lugar de requerir una medición experimental para obtener el valor del descriptor) allanó el camino para realizar predicciones de actividad biológica de compuestos hipotéticos, cuya síntesis no se hubiera logrado o intentado todavía. Más aún, hasta la incorporación de descriptores teóricos en los estudios QSAR, la predicción de la actividad implicaba la determinación experimental de descriptores empíricos para una serie de sustituyentes o compuestos de actividad conocida (los compuestos que conforman el conjunto de entrenamiento del que se deriva el modelo) y también para aquellos compuestos cuya actividad se deseaba predecir, lo que evidentemente limitaba las aplicaciones prácticas de la metodología. **La predicción de actividad involucraba, indefectiblemente, disponer de todos los compuestos predichos y realizar una o varias determinaciones experimentales** (según se incluyeran uno o varios parámetros empíricamente determinados en el modelo QSAR).

El cálculo teórico de descriptores moleculares pronto puso en evidencia la necesidad de contar con modelos o representaciones cada vez más exactos de la estructura molecular, bajo la premisa de que cuanto más exacta fuera dicha representación mejor se aproximarían los valores teóricos de los descriptores a los valores observados. Más aún: los avances en el campo de la biología establecieron simultáneamente la importancia de la conformación de la molécula para su interacción con el receptor biológico. Parecía lógico inferir que sólo una representación molecular que tuviera en cuenta la geometría más probable de la molécula en el medio biológico sería capaz de



conducir a predicciones de actividad acertadas. No obstante, como se verá más adelante, se ha demostrado que el conocimiento de la conformación más probable es en ocasiones insuficiente para predecir la capacidad de interacción con un receptor dado. En otras circunstancias el uso de representaciones más sencillas (por ejemplo, representaciones en el plano) de la estructura química implica ventajas considerables para algunas aplicaciones específicas. En la Tabla 3.2 presenta un análisis comparativo de los estudios SAR tradicionales y de las metodologías QSAR.

Tabla 3.2. Comparación de las características de las metodologías SAR clásicas y las metodologías QSAR.

SAR tradicional	QSAR
Tiene por objetivo establecer relaciones estructura-actividad sencillas para establecer cuáles descriptores o aspectos estructurales son significativos para la actividad biológica, es decir, el incremento de qué descriptor o propiedad de la molécula se asocia a un incremento de su actividad, orientando racionalmente la optimización de un compuesto líder	Logra plantear algoritmos donde se permite inferir cuali y cuantitativamente qué descriptor/es son responsables de la actividad biológica. Las aplicaciones más generales incluyen establecer hipótesis sobre los mecanismos moleculares por los que una droga interactúa con un receptor e identificar nuevos compuestos líderes a través del análisis de bases de datos de estructuras químicas
Parte de representaciones sencillas de la estructura molecular, generalmente representaciones en el plano	Parte de representaciones de complejidad variable de la estructura molecular, desde representaciones muy simples hasta muy complejas
Se utiliza para analizar series de compuestos estructuralmente similares	Puede utilizarse para analizar series similares o conjuntos de estructuras de gran diversidad
Las relaciones estructura-actividad que detecta son relativamente simples y suelen ser apreciables a simple vista por un químico experto	Las relaciones estructura actividad que detecta pueden ser muy sutiles, por cuanto los descriptores capturan información estructural no evidente
El éxito de la relación establecida depende fuertemente de la experticia del investigador	Si bien se requiere criterio químico para analizar/interpretar los resultados obtenidos, el uso de métodos estadísticos y la capacidad de descripción de los descriptores permiten lograr aplicaciones exitosas sin necesidad de una gran experticia ⁶ .

⁶ Subrayaremos de nuevo, sin embargo, que la suerte favorece a las mentes preparadas; a mayor conocimiento del observador mayor la información que podrá ser extraída de cualquier observación realizada.



Clasificación de las aproximaciones QSAR

Consideraremos dos clasificaciones de las metodologías QSAR:

- a) según la dimensionalidad de los descriptores incluidos en el modelo y;
- b) según la naturaleza de la variable dependiente considerada.

Pese a que diferentes clasificaciones de los descriptores moleculares según su dimensionalidad han sido propuestas (algunas veces con diferentes criterios de inclusión de los descriptores en cada categoría) [15-17] podemos considerar fundamentalmente dos categorías de descriptores y de aproximaciones QSAR, cada una de ellas incluyendo subcategorías cuyos límites son más difusos, como puede verse en la literatura citada:

- **Descriptores cuyo valor no depende de la conformación ni de la orientación 3D de la molécula.** Este grupo engloba varios subgrupos de descriptores moleculares tales como los descriptores 0D y 1D (descriptores constitucionales y aquellos que indican presencia o frecuencia de un grupo funcional o subestructura/fragmento determinado en la molécula) y los descriptores 2D (descriptores topológicos). Ejemplos de descriptores constitucionales son el peso molecular, el número de átomos que componen la molécula y el número de dobles o triples enlaces incluidos en la estructura. Los descriptores topológicos se discutirán en detalle en el capítulo siguiente. De aquí en más llamaremos, genéricamente, a las metodologías QSAR que utilizan descriptores de baja dimensionalidad (0D-2D), independientes de la conformación y orientación de la molécula, metodologías 2D QSAR.



- **Descriptores cuyo valor depende de la conformación y/o de la orientación 3D de la molécula.**

La metodología 3D QSAR es mucho más compleja que la aproximación 2D QSAR, requiriendo varias etapas para la obtención de los descriptores moleculares asociados. En primer lugar, debería definirse la conformación de partida en base a la cual se realiza el cálculo de los descriptores. Esta 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). **Esto implica un muestreo sistemático del espacio conformacional** (o, alternativamente, cuando esto no fuera posible por tratarse de estructuras de elevado peso molecular y gran flexibilidad, mediante otras aproximaciones como estudios estocásticos). La necesidad de obtener, en general, la conformación de mínima energía de la molécula introduce algunos interrogantes: **¿qué método debe utilizarse para calcular la energía conformacional? ¿mecánica molecular, métodos semiempíricos o métodos ab initio? ¿Es una aproximación lo suficientemente buena de lo que sucede en un medio biológico la minimización de la energía en el vacío o debe simularse un medio acuoso?**

Adicionalmente, los métodos 3D QSAR más populares, entre los que se puede destacar el CoMFA [18], se basan en disponer cada molécula (optimizada) en una grilla y calcular, posteriormente, el valor de la fuerza que diferentes campos (electrostático, estérico) ejercen sobre los vértices de la misma (a menudo utilizando un átomo o molécula de prueba), estableciendo qué campo, y en qué zona de la grilla, es significativo para la actividad biológica (fig. 3.6). Otros



ejemplos de este tipo de metodologías basadas en grilla son SOMFA [19] y GRID [20]. A finales del 2005 ya se habían publicado más de mil trabajos en revistas científicas internacionales utilizando CoMFA, la más difundida de estas técnicas [16]. Se desprende que las técnicas 3D QSAR son, además de sensibles a la conformación de la molécula que se ingresa, a la rotación y traslación de la molécula dentro de los confines de la grilla. En otras palabras, la orientación de la molécula es un aspecto crítico, y **a la fecha no se han logrado soluciones enteramente automáticas y satisfactorias a este problema cuando se trabaja con grandes colecciones de compuestos estructuralmente diversos**, requiriéndose la introducción de criterios arbitrarios por parte del operador humano. Si bien un aumento en la resolución de la grilla (reduciendo la distancia entre los puntos de la misma) conducirá en general a una mayor capacidad predictiva logrando además cierta independencia a pequeñas variaciones traslacionales y rotacionales, también producirá un crecimiento exponencial de la cantidad de variables consideradas, lo cual, como se verá, introduce dificultades adicionales debidas al incremento del costo computacional y a problemas relacionadas con el análisis estadístico de una cantidad de variables demasiado grande [21]. Para resolver estas dificultades se han introducido descriptores y técnicas dependientes de la conformación pero no de la orientación (que algunos autores refieren como descriptores 2.5D) [22, 23]. Sin embargo, los resultados de estas metodologías son más difíciles de interpretar que los de las metodologías basadas en grillas, y es por este motivo que en la mayoría de los casos los descriptores independientes de la alineación no son



utilizados fuera del grupo de investigación que los define [24].

Un análisis más profundo de los fundamentos de las metodologías 3D-QSAR nos enfrenta a una nueva problemática: cuando se asume que la conformación de mínima energía interaccionará con el receptor biológico se está realizando una simplificación extrema de la situación dinámica que acontece en la realidad. Los sistemas biológicos son sistemas químicos complejos y dinámicos y no existen garantías de que la conformación de mínima energía corresponda a la conformación bioactiva: más aún, se sabe que tanto el ligando como el receptor inducen cambios conformacionales el uno en el otro, recíprocamente, siendo éste el disparador del evento biológico. Si bien la simplificación de considerar una molécula rígida interaccionando con un receptor rígido ha resultado exitosa en muchos estudios, en particular en aquellos casos en los que el ligando es una molécula rígida con la libre rotación de los enlaces restringida, los efectos dinámicos son en muchos casos considerables; este hecho ha derivado en la generación de metodologías 4D-QSAR que muestrean no una sino varias orientaciones, conformaciones, estados de protonación y esteroisómeros de los ligandos estudiados [25-27]. Recientemente ha sido introducido el paradigma 5D-QSAR, que tiene en cuenta no sólo la flexibilidad conformacional del ligando sino también del receptor [28]. Estas metodologías están en etapa incipiente y resta evaluar si el creciente costo computacional requerido es compensado por un incremento equivalente en la capacidad predictiva de los modelos 4D y 5D-QSAR.



Según la naturaleza de la variable dependiente considerada, los modelos QSAR pueden también dividirse en cualitativos o cuantitativos. La primera clase (que también podemos referir como modelos clasificatorios o análisis discriminante) considera una variable dependiente discreta, categórica; la propiedad observada es representada mediante una etiqueta (tóxico y no tóxico; activo, moderadamente activo e inactivo; etc.). Los modelos cuantitativos, en cambio, utilizan una variable dependiente continua y se usan para predecir, por lo tanto, el valor numérico de una propiedad determinada (por ej. ED_{50}).

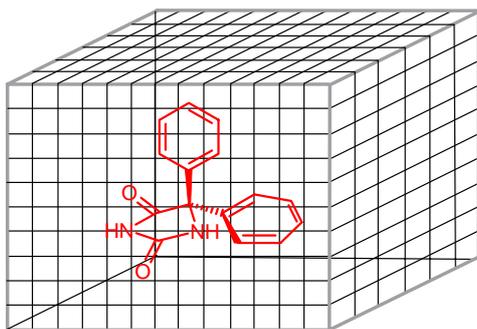


Fig. 3.6. Las metodologías 3D-QSAR basadas en grillas (dependientes de la conformación y orientación de la molécula en la grilla). Los descriptores surgen de calcular el valor de un campo determinado (por ej. un campo coulombico para las interacciones electrostáticas) sobre un átomo o molécula de prueba ubicado en los vértices de la grilla. En el ejemplo, una representación de una grilla en cuyo interior se encuentra una molécula de difenilhidantoína.

Fundamentos de la metodología elegida para la presente tesis.

En el presente trabajo se ha decidido por la búsqueda de nuevos agentes antiepilépticos **mediante 2D-QSAR basado en descriptores y enfocado desde el ligando**. Esta decisión supone algunas ventajas y desventajas respecto a otras alternativas posibles. Entre las ventajas podemos mencionar:

- a) la independencia del valor de los descriptores 2D respecto de la conformación y orientación de las estructuras cuyo valor ha de predecirse, lo cual reduce considerablemente el trabajo de pre-



procesamiento de las estructuras de las bases de datos (no se requieren coordenadas optimizadas ni establecer criterios de orientación para el cálculo de los descriptores). Dado que se desea analizar bases de datos de 10^5 a 10^6 estructuras con gran diversidad química, el solo pre-procesamiento de las estructuras cuya actividad biológica se busca predecir demandaría un tiempo considerable e implicaría definir protocolos de optimización y criterios de orientación por parte del operador humano, sobre todo si se tiene en cuenta que se desean identificar estructuras bloqueantes de canales de sodio voltaje operados. Se trata de un tipo de receptor particularmente complejo que consta de una subunidad α formada por cuatro dominios (cada uno de ellos consta de seis α -hélices transmembrana) y una o más subunidades β de peso molecular relativamente menor (fig. 3.3); los canales de sodio poseen múltiples sitios de unión a inhibidores competitivos y no competitivos y su estructura terciaria no ha sido aún enteramente dilucidada; más aún, los diferentes subtipos de canales de sodio del SNC presentan diferente afinidad por los fármacos [29,30]. Todos estos hechos, asociados a que no se dispone de una estructura de rayos X del canal de sodio, dificultan utilizar el receptor para definir una orientación preferencial de moléculas estructuralmente diversas en la grilla.

- b) el bajo costo computacional que supone el cálculo de descriptores 2D, lo cual facilita el análisis de grandes colecciones de estructuras virtuales en un tiempo reducido;
- c) no se requiere información sobre la estructura del receptor.

Desde luego, la metodología elegida también tiene sus limitaciones. Las metodologías 3D-QSAR permiten generar modelos gráficos que permiten una fácil interpretación de los resultados observados,



posibilitando la identificación de las interacciones fármaco-receptor más importantes para la actividad biológica. Sin embargo, nada dicen respecto a otros procesos muy importantes para que un compuesto sea activo, tales como el pasaje a través de las distintas barreras biológicas, la metabolización, la toxicidad, etc. Los descriptores 2D son en cambio bastante abstractos y codifican/condensan, en un solo número, información respecto a diversos aspectos estructurales. Si bien son muy eficaces para identificar nuevos líderes, es en general difícil extraer de un modelo 2D QSAR información significativa desde el punto de vista fisicoquímico acerca de los mecanismos por los cuales una droga determinada interacciona con su receptor molecular. Los descriptores 2D funcionan, efectivamente, como una caja negra de la que se extrae información útil pero difícil de interpretar. Por ejemplo, si un modelo 2D-QSAR predice que un compuesto determinado presentará baja actividad biológica, será difícil definir si esa baja actividad predicha se debe a su baja biodisponibilidad, su rápida eliminación o a que no interacciona efectivamente con el blanco molecular (por más que la predicción sea acertada).

A modo de síntesis se podría decir que las metodologías 3D describen con mayor claridad la interacción de la droga y su receptor, evento necesario pero no suficiente para que un fármaco ejerza su actividad (los otros eventos deberán ser modelados, se desprende, mediante modelos complementarios); adicionalmente, requieren un pre-procesamiento de la estructura que no las hace aptas para el estudio de grandes bases de datos. Las metodologías 2D QSAR, si bien por su carácter abstracto y holístico no arrojan tanta luz sobre el proceso de unión al receptor, capturan otros aspectos necesarios para producir la bioactividad, y por otro lado la velocidad de cálculo de descriptores de baja dimensionalidad y la independencia de la conformación y orientación de la molécula las hace ideales para



identificar nuevos líderes en grandes bases de datos de diversidad estructural.

En el trabajo de tesis no sólo se han generado modelos 2D QSAR para predecir actividad anticonvulsiva por bloqueo de canales de sodio voltaje-operados, sino también, en forma complementaria, filtros particulares para predecir la biodisponibilidad a nivel del SNC y la neurotoxicidad de los compuestos evaluados. Puede subrayarse que, dado al carácter holístico de los descriptores topológicos, una adecuada selección del conjunto de entrenamiento utilizado contribuye a seleccionar compuestos biodisponibles y seguros durante el proceso de SV. Sin embargo, hemos decidido refinar esta selección mediante generación de filtros específicos de absorción y neurotoxicidad. En la

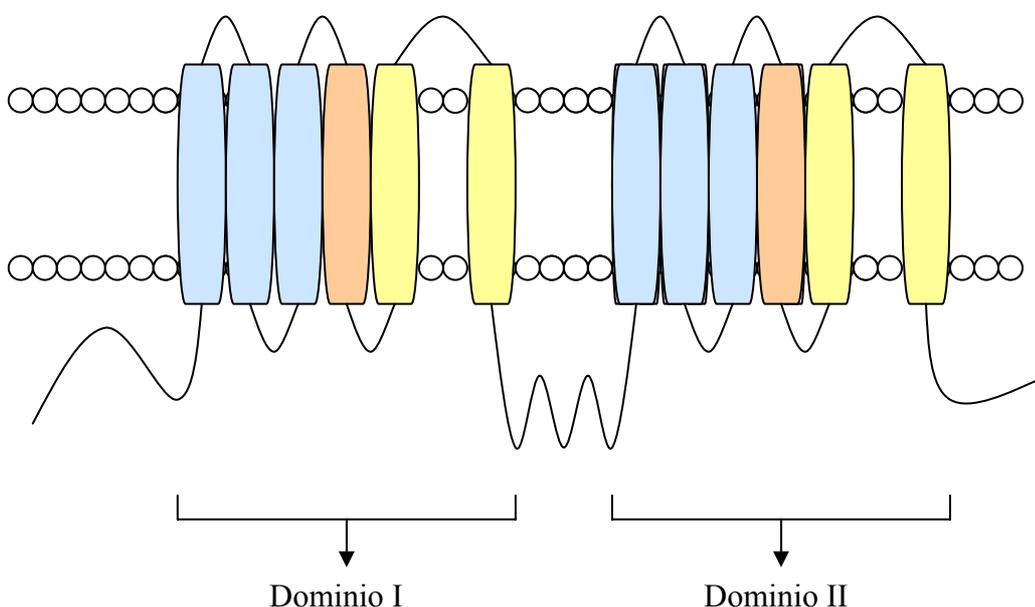


Fig. 3.3. Representación esquemática de la estructura terciaria de la subunidad α de un canal de sodio voltaje operado. Se muestran dos de los cuatro dominios de la misma; cada dominio está formado por seis α -hélice transmembrana. En anaranjado, el segmento que reconoce los cambios de voltaje; en amarillo, los segmentos que definen el poro por el que se transporta el sodio. La subunidad α puede se asocia a una o más subunidades β . Existen numerosos sitios de unión a fármacos intra y extracelulares.



fig. 3.4 se esquematizan distintos procesos que condicionan la actividad de un fármaco que actúa a nivel del sistema nervioso central, y cuáles de ellos y en qué forma se han considerado en la estrategia QSAR elegida. Hemos elegido para el esquema la vía de administración oral, forma farmacéutica sólida, que por varias razones es considerada en general como la vía de preferencia para medicamentos que se administran en forma crónica (entre ellas: inocuidad, estabilidad, facilidad de dosificación y administración y facilidad de transporte de las formas sólidas orales).

Etapas generales de un estudio QSPR/QSAR

En general, un estudio QSAR pueden dividirse en las siguientes etapas:

- 1. Generación del conjunto de entrenamiento.** Se denomina conjunto de entrenamiento al conjunto de estructuras químicas a partir de las cuales se deriva el modelo QSAR. El conjunto QSAR define el “espacio químico” dentro del cual la aplicación del modelo tendrá validez. En otras palabras, el valor de la actividad/propiedad biológica predicha para un compuesto externo al conjunto de entrenamiento será válida si y sólo si el mismo posee cierta similitud con alguna o varias de las estructuras incluidas en el conjunto de entrenamiento. El espacio químico dentro del cual la aplicación del modelo QSAR es confiable se denomina “dominio de aplicabilidad”. La forma más sencilla de estimar el dominio de aplicabilidad cuando los valores de los descriptores incluidos en el modelo están bien distribuidos para el conjunto de entrenamiento, es definir un espacio N-dimensional delimitado por los rangos de valores que los N descriptores del modelo asumen para el conjunto de entrenamiento [31,32]. Otra regla útil a tener en cuenta cuando se diseña un



conjunto de entrenamiento es que los valores observados de la propiedad estudiada deberían presentar una buena distribución (que aproxime una distribución normal) en un rango de entre tres y cuatro órdenes logarítmicos. Este requisito posee un fundamento lógico: supongamos un conjunto de entrenamiento formado solamente por compuestos de alta y mediana actividad, el modelo QSAR derivado del mismo sólo podrá predecir con exactitud valores de compuestos muy activos y moderadamente activos, pero no servirá para diferenciar compuestos activos de inactivos (no identificará qué características estructurales son francamente desfavorables a la actividad por no estar los compuestos poco activos e inactivos representados dentro del conjunto de entrenamiento del que se deriva el modelo). Un último requisito esencial es que los compuestos del conjunto de entrenamiento posean un mecanismo de acción común, esto es, interaccionen con el mismo receptor o blanco molecular. Supongamos, por ejemplo, que se desea generar un modelo QSAR para la búsqueda de compuestos antihipertensivos. Si se incluyeran en el conjunto de entrenamiento diuréticos, inhibidores de la enzima convertora de angiotensina y vasodilatadores periféricos, sería difícil derivar un único modelo QSAR que identificara las características favorables a la actividad antihipertensiva que no sean simplemente las características típicas de los fármacos (características *drug-like*, que se discutirán luego), debido a que un fármaco inhibidor de la enzima convertora de angiotensina interactúa con un blanco molecular que en principio nada tiene que ver con los varios blancos moleculares de los distintos diuréticos, de lo que se deduce que es poco probable encontrar patrones estructurales comunes relacionados con la interacción fármaco-receptor.



2. **Cálculo del valor de un conjunto de D descriptores para el conjunto de entrenamiento.** Lo fundamental en esta etapa es chequear las estructuras del conjunto de entrenamiento y, en los casos en que sea necesario (3D-QSAR en adelante) realizar un pre-procesamiento adecuado de las estructuras de los compuestos.
3. **Selección de un conjunto d de descriptores de entre los D descriptores disponibles; generación del modelo.** La Regresión Lineal Múltiple (Multiple Linear Regression - MLR) es el método tradicional para generar modelos QSAR. El mismo se basa en el método de los cuadrados mínimos, esto es, encontrar la combinación lineal de variables que minimice la suma de los cuadrados de los residuales (diferencia entre los valores observados de la variable dependiente y los valores predichos) [33]. Usualmente se utilizan el coeficiente de determinación r^2 y la desviación estándar en la estimación SEE (Standard Deviation on the Estimate) para evaluar la bondad del ajuste de los valores de la variable independiente predichos por el modelo con respecto a los valores observados. El coeficiente de determinación indica la fracción de la varianza total de los datos experimentales que puede ser explicado por el conjunto de descriptores incluidos en el modelo.

$$r^2 = 1 - \frac{\sum_{i=1}^n (Y_{obs,i} - Y_{pre,i})^2}{\sum_{i=1}^n (Y_{obs,i} - Y_{med})^2} = \frac{\sum_{i=1}^n (Y_{pre,i} - Y_{med})^2}{\sum_{i=1}^n (Y_{obs,i} - Y_{med})^2}$$

donde $Y_{obs,i}$ representa el valor observado de la variable dependiente para el caso i ; $Y_{pre,i}$ representa el valor predicho para el caso i y; Y_{med} representa el valor medio observado de la variable dependiente para el conjunto de entrenamiento y n es el



número de casos (compuestos) que integran el conjunto de entrenamiento.

SEE está definido por:

$$SEE = \sqrt{\frac{\sum_{i=1}^n (Y_{obs} - Y_{pre})^2}{n - d - 1}}$$

donde d es el número de descriptores incluidos en el modelo.

En todo análisis de MLR debe verificarse que:

- Los residuales estén normalmente distribuidos alrededor de una media de 0; presenten homoscedasticidad (es decir, varianza constante en todo el rango abarcado por los casos) y; deben ser mutuamente independientes.
- El número de variables independientes sea menor que el número de observaciones (hipótesis de rango completo).
- No existe correlación lineal significativa entre las variables independientes. En general, se pone una cota al coeficiente de correlación admitido para cualquiera de los pares de descriptores incluidos en el modelo. La cota puede ser, según el operador, de 0.5 (acorde al criterio más conservador), 0.7, 0.9 ó 0.95. **Se observa que los criterios menos conservadores violan una de las hipótesis del análisis de regresión lineal, además de dificultar seriamente la interpretación del modelo generado.** Otra forma de restringir el ingreso de pares de descriptores altamente correlacionados en el modelo permitida por los programas de estadística disponibles es fijar una cota a la tolerancia β definida como $1 - r^2$, donde r^2 es en este caso el coeficiente de determinación entre una de las

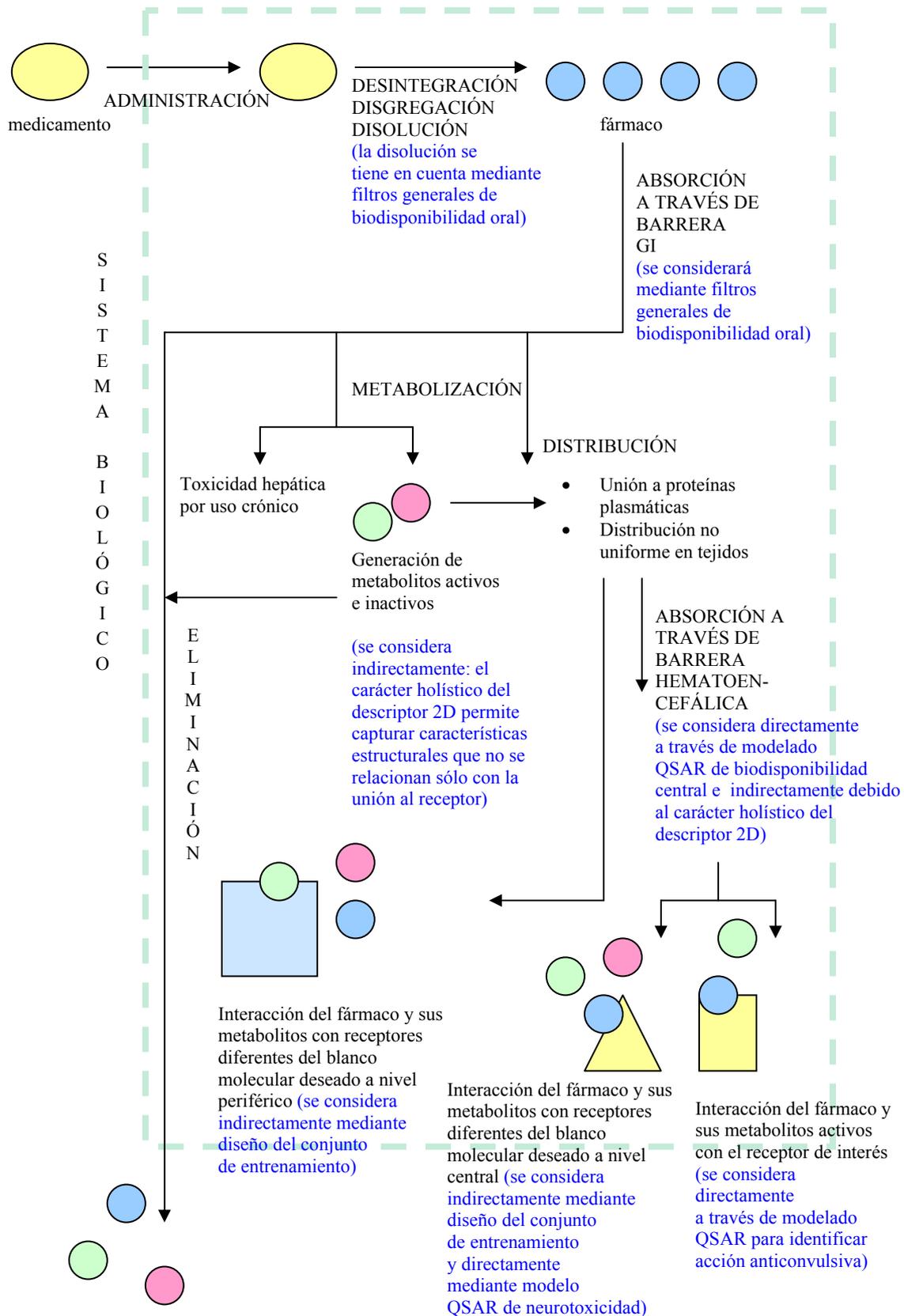


Fig. 3.4. Esquema de los diferentes procesos que el medicamento y el fármaco atraviesan en el organismo; en azul, se consigna cómo se consideran los diferentes procesos en el presente trabajo.



variables independientes del modelo y todas las demás variables independientes. Sin embargo, cabe mencionar que Randic ha demostrado que la inclusión de pares de variables altamente correlacionadas **con signo opuesto** puede ser estadísticamente significativa [34,35].

Las ventajas de la MLR son su simplicidad de cálculo y la posibilidad de interpretar los resultados obtenidos con relativa facilidad. El principio de parsimonia indica elegir el modelo más simple entre aquellos que explican adecuadamente la problemática estudiada. **Los motivos para preferir modelos más simples son prácticos además de filosóficos: adicionar variables inútiles implica que, para aplicar el modelo se deberá calcular el valor de esos descriptores no significativos, con el consabido desperdicio de tiempo y recursos humanos; incluir predictores innecesarios en un modelo usualmente conduce a decisiones equivocadas** (por ejemplo, si se decide incluir, como descriptor, en un modelo QSAR, el número de grupos -OH y este predictor es en realidad irrelevante, al aplicar el modelo se descartarán o elegirán compuestos según su número de grupos -OH, pudiendo perder líderes valiosos o reteniendo compuestos sin valor); **adicionar variables irrelevantes introduce variación azarosa -ruido- en el modelo; finalmente, cuanto más simple el modelo mayor su portabilidad.** Sin embargo, las razones filosóficas también son importantes: los modelos más simples son más fáciles de interpretar y son considerados más “bellos” en el sentido de la estética matemática y científica en general. Por ejemplo, un modelo de un descriptor será preferible a uno de diez descriptores que posea un desempeño similar. Las técnicas de selección de variables independientes más utilizadas en conjunto con la MLR son las conocidas como



Stepwise Forward y Stepwise Backward. La metodología Stepwise Forward consiste en llevar a cabo una regresión por cuadrados mínimos por pasos. En cada paso o etapa la variable independiente del pool de D variables mejor correlacionada con la dependiente es incluida en la ecuación de regresión; en el segundo paso, se adiciona la variable independiente que de entre todas las remanentes produce un incremento mayor en el r^2 ; este proceso continúa hasta que la adición de una nueva variable no contribuye a un aumento estadísticamente significativo del r^2 . Un test de F parcial puede ser utilizado para corroborar si la inclusión de una variable en un modelo induce un aumento significativo de r^2 . El estadístico F se define como:

$$F = \frac{r^2/d}{(1-r^2)/(n-k-1)} = \frac{(n-d-1)}{1-r^2} \left(\frac{r^2}{d} \right)$$

donde se observa que la significancia estadística aumenta lógicamente con el número de casos y el valor de r^2 , pero es inversamente proporcional al número de descriptores incluidos en el modelo y tiende a 0 cuando d se aproxima a n (esto es, a medida que aumenta el número de variables incluidas en el modelo aumenta la probabilidad de correlación azarosa, es decir, que fortuitamente la varianza de alguna de las variables independientes remanentes en el pool refleje la varianza aún no explicada por el modelo; el hecho de que al ser d igual a n la significancia estadística se vuelva nula refleja la hipótesis de rango completo de la que parte el análisis de regresión). Alternativamente, la metodología Stepwise Backward comienza incluyendo todas las variables del pool en el modelo y elimina, una a una, variables independientes hasta que la eliminación de



una variable reduce significativamente el r^2 . Otros métodos lineales populares además de MLR son el análisis de componentes principales (Principal Component Analysis - PCA) y el método de los cuadrados parciales mínimos (Partial Least Squares - PLS). Ambos métodos son útiles para reducir la dimensionalidad del pool de descriptores cuando el número total de descriptores D es mucho mayor que el número de casos n (sin realizar una selección explícita), seleccionando como variables independientes ciertas combinaciones lineales y ortogonales de los mismos [33,36]. En el caso de PCA esas variables artificiales se denominan componentes principales, y se obtienen utilizando la covarianza entre los distintos descriptores como criterio de ponderación, pero sin tener en cuenta la covarianza entre los valores observados de la variable dependiente y los descriptores. En el caso de PLS, las variables artificiales se denominan variables latentes y en este caso sí se tiene en cuenta la covarianza entre la variable dependiente y los descriptores para ponderar las distintas variables. En otras palabras, no hay garantía del valor predictivo que tendrán las componentes principales, mientras que sí la hay en el caso de las variables latentes de PLS.

Existen otros métodos no lineales de selección de variables y generación de modelos, entre los que encontramos los algoritmos genéticos (Genetic Algorithms - GA) y las redes neuronales (Neural Networks - NN); ambos métodos han sido inspirados por la naturaleza: en el primero se toma como modelo la teoría de la evolución y la selección natural para seleccionar los “modelos más aptos”; el segundo se inspira en la capacidad del cerebro animal para procesar información de manera no lineal [37].

En el caso de los GA, un conjunto de representaciones abstractas (llamados cromosomas) de una población de posibles soluciones



(llamadas individuos) se aparean o recombinan. Los genes de cada cromosoma representan la presencia o ausencia de un variable independiente determinada en la solución/individuo. Cada N generaciones (N definido por el operador humano) se evalúa el desempeño de los modelos que conforman la descendencia, seleccionándose los más aptos para nuevas recombinaciones (en algunos casos se utiliza una selección estocástica, generando distintos grupos de soluciones según su “aptitud” y seleccionando para la siguiente serie de recombinaciones individuos de todos los grupos, pero en mayor proporción del grupo más apto). Cada tantas generaciones se suelen introducir, para imitar a la naturaleza, mutaciones (deleciones de genes, inserciones, etc.) y se pueden incluir operadores que produzcan, al azar, la muerte de algún individuo determinado con la extinción del genotipo correspondiente (ver fig. 3.5). En el caso de las llamadas estrategias evolutivas (ES), al contrario de lo que ocurre con los GA, la frecuencia de mutaciones predomina sobre la de recombinaciones, y en lugar de seleccionarse estocásticamente un grupo de soluciones que incluye individuos aptos y poco aptos, se tiende a seleccionar sólo al individuo más apto para dar origen a la siguiente generación de mutantes [38, 39].

En ocasiones la interdependencia entre un proceso biológico y alguna propiedad derivada de la estructura molecular es no lineal (un ejemplo clásico es cómo suele variar la actividad biológica según el coeficiente de partición octanol-agua de las moléculas). Las redes neuronales tratan de imitar, de forma simplificada, la estructura de las unidades de procesamiento del cerebro, las cuales, con una velocidad de cálculo ciertamente muy inferior a la de las supercomputadoras



actuales logran realizar tareas muy complejas, por ejemplo en lo relativo al procesamiento de estímulos visuales y auditivos. Esquemáticamente, el cerebro humano está formado por diez mil millones de neuronas interconectadas masivamente (con un promedio de varios miles de interconexiones por neurona). Cada neurona es una célula especializada que puede propagar una señal eléctrica, con una estructura ramificada por la que ingresa la información (las dendritas) y otra estructura ramificada por la que la información se propaga a otras neuronas (el axón). Cuando la neurona recibe un estímulo por encima de cierto umbral, se

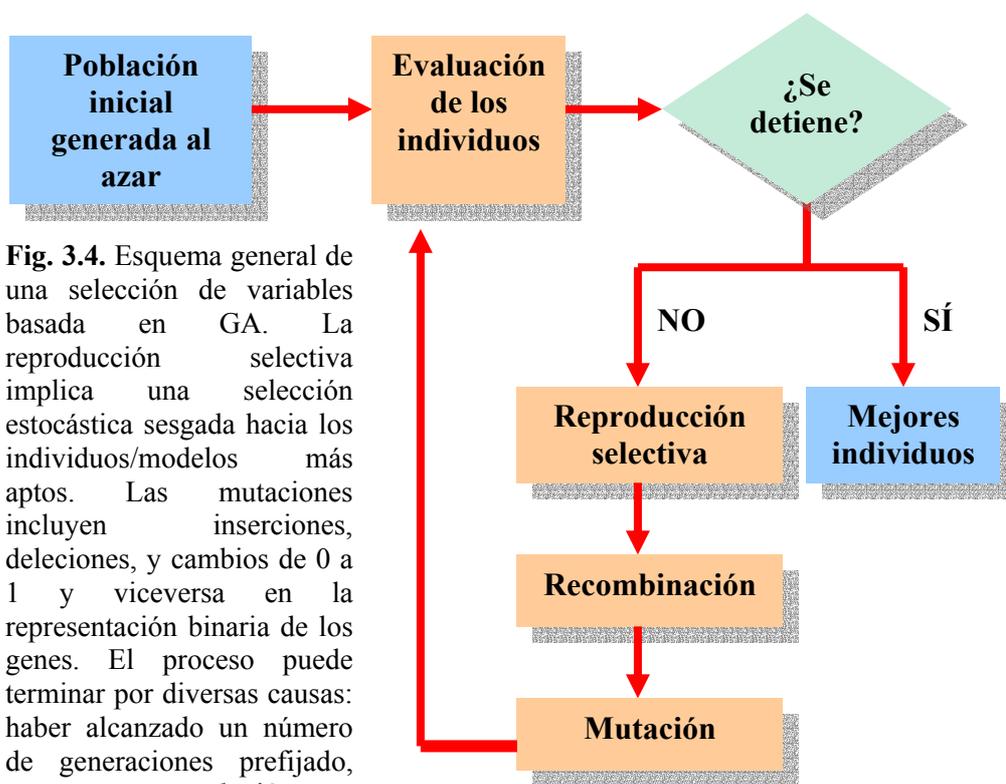


Fig. 3.4. Esquema general de una selección de variables basada en GA. La reproducción selectiva implica una selección estocástica sesgada hacia los individuos/modelos más aptos. Las mutaciones incluyen inserciones, deleciones, y cambios de 0 a 1 y viceversa en la representación binaria de los genes. El proceso puede terminar por diversas causas: haber alcanzado un número de generaciones prefijado, encontrar una solución que se considere aceptable, etc.

activa y envía una señal neuroquímica a otras neuronas, a través del axón. El aprendizaje consiste, fundamentalmente, en alterar la fuerza de ciertas conexiones sinápticas, favoreciéndolas. En síntesis, a partir de un gran número de unidades de procesamiento relativamente sencillas (cada una realizando una suma ponderada de sus estímulos y disparando una señal binaria



si la misma supera cierto umbral) el cerebro logra realizar tareas sumamente complejas. Las redes neuronales imitan este proceso: reciben y ponderan y transforman estímulos (*input*, el valor de las variables independientes) y generan respuestas (*output*, la predicción de la categoría a la que pertenece una molécula o el valor de una propiedad fisicoquímica determinada de la misma). En la fig. 3.6 se presenta una arquitectura típica de una red neuronal, consistente en una capa de neuronas de entrada (que no realiza otra operación más que distribuir el vector de datos de entrada -vector formado por las variables independientes- a la siguiente capa) una capa de neuronas escondida y una capa de de neuronas de salida.

Típicamente, las neuronas de las capas escondida y de salida transforman los datos que proceden de la capa anterior mediante funciones sigmoideas de la forma:

$$\text{Sigm}(\text{datos de entrada}) = \frac{1}{1 + e^{-(\text{datos de entrada})}}$$

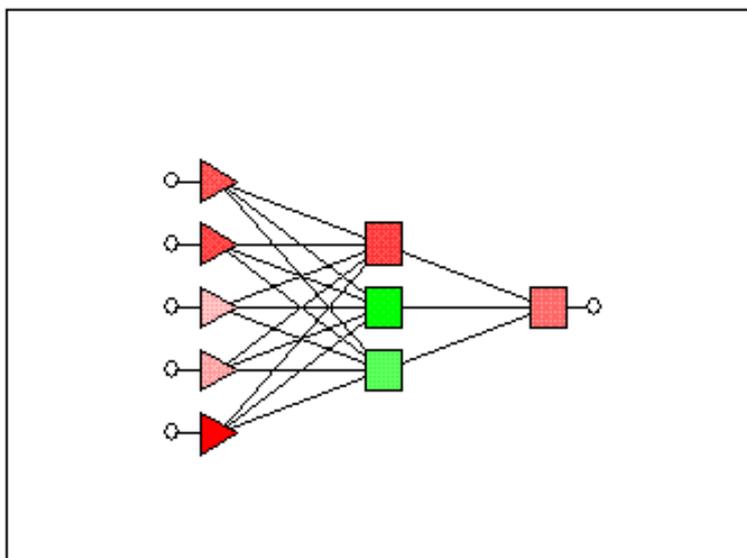


Fig. 3.6. Ejemplo de una arquitectura de red neuronal, con cinco neuronas en la capa de entrada correspondientes a cinco variables independientes, tres neuronas en la capa escondida y una neurona en la capa de salida.



En la red neuronal del ejemplo, la transformación final que realizaría la única neurona de la capa de salida sería:

$$f(x) = \text{Sigm}\left(\sum_{j=1}^3 v_j \text{Sigm}\left(\sum_{i=1}^5 \omega_{i,j} x_i - \theta_j\right) - \theta^{\text{salida}}\right)$$

donde θ_j y θ^{salida} son los valores umbral de cada una de las neuronas de la capa escondida y de la capa de salida, respectivamente, x_i representa a las variables independientes consideradas, $\omega_{i,j}$ son los coeficientes con los que cada neurona de la capa escondida pondera las variables independientes y v_j representa a los coeficientes con los que la capa de salida pondera la información procedente de cada una de las neuronas de la capa escondida.

La ventaja de las redes neuronales es, claramente, la posibilidad de modelar funciones complejas; su desventaja es que para entrenar una red neuronal se requieren, en general, entre cientos y miles de casos, muchos más de los necesarios para generar un modelo lineal (en el caso de las aplicaciones en Química Medicinal el número de compuestos activos conocidos es limitado). La capacidad de generar modelos complejos acentúa la tendencia a la sobreparametrización, es decir, que el modelo “memorice” las características del conjunto de entrenamiento pero no tenga capacidad predictiva sobre casos no utilizados en el entrenamiento. Por lo tanto, se suele reservar, además de un conjunto de prueba, un conjunto de verificación utilizado para chequear, durante el entrenamiento, que no se esté sobreparametrizando. La necesidad de contar con tres conjuntos de casos independientes y representativos disminuye considerablemente el número de casos disponibles para entrenar la red. Por otro lado, no se trata de un método de selección de variables, sino de generación de modelos, por lo que se debe



tener de antemano cierto conocimiento teórico o intuitivo sobre cuáles de entre el conjunto total de variables son las que contribuyen más a la propiedad biológica estudiada, o de lo contrario combinar las redes neuronales con métodos de selección de variables como los GA [33, 37].

Las consideraciones que deben tenerse en cuenta en la etapa de selección de descriptores y generación del modelo son varias. En primer lugar, **se ha demostrado que el uso de un conjunto total de descriptores D grande incrementa ostensiblemente la probabilidad de encontrar correlaciones azarosas entre un subconjunto de descriptores d y la variable dependiente** [40].

Cuando el pool de descriptores es grande en comparación con el número de casos deberá, por lo tanto, ponerse énfasis en el proceso de validación del modelo. En nuestro caso se utilizará un pool de descriptores grande, pero el casi nulo pre-procesamiento de las estructuras y la facilidad de cálculo característicos de los descriptores 2D facilitan el proceso de validación, minimizando la probabilidad de correlación fortuita.

Una problemática extra a tener en cuenta es la cuestión, ya mencionada fugazmente, de la sobreparametrización, esto es, la violación del principio de parsimonia, conocido también como la navaja de Occam¹: incluir más parámetros o utilizar modelos más flexibles de lo necesario (por ejemplo, una ecuación polinómica o una red neuronal pueden ajustar los valores de los compuestos del conjunto de entrenamiento mediante una relación curvilínea, no lineal, no sólo sin introducir beneficios sino a expensas del desempeño del modelo sobre un conjunto externo). Hawkins ha publicado un trabajo detallado sobre los riesgos de la

¹ Entia non sunt multiplicanda praeter necessitatem. “No ha de presumirse la existencia de más cosas que las absolutamente necesarias”. La explicación más simple y suficiente es la más probable, aunque no siempre la verdadera. Con la enunciación de este principio, el fraile inglés Guillermo de Occam (u Ockham) expresaba, en el siglo XIV, la ruptura entre el pensamiento científico y el pensamiento teológico o mágico.



sobreparametrización y la utilidad de las herramientas de validación para salvaguardarse de la misma [41]. Una regla empírica útil para resguardarse de la sobreparametrización es vigilar que la relación n/d sea mayor a cinco.

- 3. Validación del modelo.** El proceso de validación es una etapa crítica del modelado QSAR. Tiene por objeto la determinación de la robustez y capacidad predictiva del modelo [42]. La validación puede ser interna y/o externa. **En el caso de la validación interna, se reutiliza el mismo conjunto de entrenamiento para validar el modelo generado mediante validación cruzada y ensayo de aleatorización de Fisher.** Los procedimientos de validación cruzada de modelos QSAR por excelencia son las técnicas denominadas Leave-One-Out (LOO - “Dejar uno afuera”) y Leave-Group-Out (LGO - “Dejar un grupo afuera”, también llamada por algunos autores Leave-Some-Out o Leave-N-Out). Consisten en, una vez generado el modelo, remover grupos de m compuestos del conjunto de entrenamiento, recalculando el modelo con los compuestos remanentes y utilizar el nuevo modelo para predecir la actividad de las estructuras removidas. El caso extremo en el que m es igual a uno constituye la metodología LOO (“dejar uno afuera”). Los parámetros que se utilizan para estimar la capacidad predictiva del modelo mediante esta técnica son análogos a los que se describieron en la etapa de generación del modelo, $SEE_{LOO/LGO}$ y r^2_{cv} , o más comúnmente, q^2 :

$$q^2 = 1 - \frac{\sum_{i=1}^n (Y_{obs,i} - Y_{pre,i})^2}{\sum_{i=1}^n (Y_{obs,i} - Y_{med})^2}$$



donde ahora $Y_{pre, i}$ corresponde al valor predicho para cada compuesto i cuando es removido del conjunto de entrenamiento.

Aunque en general se considera aceptable un valor de q^2 mayor a 0.5, se ha demostrado que un valor adecuado de q^2 es condición necesaria pero no suficiente para afirmar la capacidad predictiva del modelo [43]. El q^2 tiende a ser sobreoptimista, en especial en aquellos casos en los que el conjunto de entrenamiento no es estructuralmente diverso (por ejemplo, cuando está formado por derivados más o menos similares de un mismo compuesto líder, lo cual no es del todo inusual en el campo de la Química Medicinal) [41]. Adicionalmente, dado que al contrario del r^2 el q^2 no aumenta indefinidamente a medida que se adicionan parámetros al modelo y por el contrario alcanza un máximo, se utiliza en algunos casos (por ej., en PLS) para determinar el número óptimo de variables que se incluirán en el modelo.

La aleatorización de Fisher, destinada a minimizar la probabilidad de correlación fortuita entre los descriptores y la variable dependiente, se lleva a cabo “mezclando” al azar los n valores observados de la variable dependiente entre los n compuestos, regenerando el modelo con la variable dependiente aleatorizada. Este procedimiento se suele repetir n veces. Idealmente, el modelo original será estadísticamente muy superior a los modelos aleatorizados.

Por último, la validación externa consiste en predecir el valor de la variable dependiente para un conjunto de compuestos cuya actividad se conoce y que no han sido utilizados para generar el modelo. El principal problema de este proceso es que deben resignarse casos que podrían utilizarse en la generación del modelo para evaluar su predictividad, lo cual no es viable cuando se conocen pocas estructuras con valor experimental



conocido de la propiedad biológica de interés. Las opiniones, respecto a lo conveniente en este punto, están divididas. Algunos autores afirman (y han realizado algunas demostraciones al respecto) que el uso de un conjunto de prueba es sólo admisible cuando se dispone de un conjunto de entrenamiento muy grande (cientos de compuestos) [41]. Aparentemente, apartar un conjunto de prueba cuando no se dispone de cientos de compuestos atentaría contra la predictividad del modelo. Hawkins afirma que el conjunto de prueba como herramienta de estimación de la predictividad es válido cuando está formado por más de 20 estructuras, e, idealmente, cincuenta. Sin embargo, otros autores sugieren que entre un 10 y 20% del total de estructuras disponible es suficiente [42], y la gran mayoría de los especialistas reconoce a la validación externa como etapa imprescindible [43]. En nuestro caso hemos considerado una situación intermedia, validando siempre interna y externamente todos los modelos generados durante la tesis y reservando, para la validación externa, no un número fijo de estructuras sino un porcentaje del número de compuestos incluidos en el conjunto de entrenamiento, no menor al 30% del mismo.

- 5. Interpretación y aplicación del modelo.** La interpretación se refiere a calcular la contribución relativa de los descriptores al modelo (esto puede hacerse fácilmente comparando los coeficientes de regresión normalizados, multiplicando el coeficiente de regresión por la desviación estándar del predictor sobre la desviación estándar de la variable dependiente o, alternativamente, normalizando los predictores en forma previa al análisis de regresión, sustrayendo la media del predictor y dividiendo por su desviación estándar; una forma adicional es multiplicar el coeficiente de regresión por el valor promedio del descriptor para el conjunto de



entrenamiento y dividir este valor por la sumatoria de los coeficientes por el valor medio de los descriptores asociados). El estudio de la definición del descriptor orientará acerca del significado fisicoquímico del modelo; en nuestro caso, ya hemos comentado que el carácter abstracto de los descriptores topológicos (ver capítulo siguiente para más detalles) complica la interpretación, aunque intentaremos de todas maneras abordar los significados de los descriptores incluidos en cada modelo. Sin embargo, esta no es una etapa crítica de este trabajo de tesis por cuanto el objetivo principal es la identificación de nuevos líderes y no el diseño de nuevas estructuras ni la elucidación de mecanismos moleculares. En la etapa de aplicación del modelo deberá tenerse presente, sobre todo, el dominio de aplicabilidad del mismo. Desde luego, la verificación experimental (mediante ensayos biológicos en el caso de predecir bioactividades) constituye la prueba definitiva de la validez del modelo.

Referencias

1. Braña, M.F.; Sánchez-Migallón, A. Anticancer drug discovery and pharmaceutical chemistry: a history. *Clin. Transl. Oncol.* **2006**, *8*, págs. 717-728
2. Nair, R.; Chabhadiya, R.; Chanda, S. Marine algae: screening for a potent antibacterial agent. *J. Herb. Pharmacother.* **2007**, *7*, págs. 73-86
3. Tulp, M.; Bohlin, L. Unconventional natural sources for future drug discovery. *Drug Discov. Today.* **2004**, *9*, págs. 450-458
4. Mayer, A.M.S.; Rodríguez, A.D.; Berlinck, M.G.S.; Hamann, M.T. Marine pharmacology in 2003-4: Marine compounds with anthelmintic antibacterial, anticoagulant, antifungal, anti-inflammatory, antimalarial, antiplatelet, antiprotozoal, antituberculosis, and antiviral activities; affecting the cardiovascular, immune and nervous systems, and other miscellaneous mechanisms of action. *Comp. Biochem. Phys. C.* **2007**, *145*, págs. 553-581
5. Patrick, G.L. An introduction to Medicinal Chemistry. Oxford University Press, New York, Estados Unidos, **2001**



6. Domino, E.F. History of modern psychopharmacology: a personal view with an emphasis on antidepressants. *Psychosom. Med.* **1999**, 61, págs. 591-598
7. Moreira Lima, L.; Barreiro, E.J. Bioisosterism: a useful strategy for molecular optimization and drug design. *Curr. Med. Chem.* **2005**, 12, págs. 23-49
8. Thornber, C.W. Isosterism and molecular modification in drug design. *Chem. Soc. Rev.* **1979**, 8, págs. 563-580
9. Patani, J.A.; LaVoie, E.J. Bioisosterism: a rational approach in drug design. *Chem. Rev.* **1996**, 96, págs. 3147-3176
10. Stánczak, A.; Ferrá, A. Prodrugs and soft drugs. *Pharmacol. Rep.* **2006**, 58, págs. 599-613
11. Koppitz, M.; Eis, K. Automated Medicinal Chemistry. *Drug Discov. Today.* **2006**, 11, págs. 561-568
12. Ganesan, A. Recent developments in combinatorial organic synthesis. *Drug Discov. Today.* **2002**, 7, págs. 47-55
13. Hansch, C.; Muir, R.M.; Fujita, T.; Maloney, P.; Geiger, E.; 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, págs. 2817-2824
14. Leo, A.; Panthananickal, A.; Hansch, C.; Theiss, J.; Shimkin, M.; Andrews, A.W. A comparison of mutagenic and carcinogenic activities of aniline mustards. *J. Med. Chem.* **1981**, 24, 859-864
15. Dudek, A.Z.; Arodz, T.; Gálvez, J. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Comb. Chem. High. T. Scr.* **2006**, 9, págs. 213-228
16. Korhonen, S. FLUFF-BALL, a fuzzy superposition and QSAR technique. Kuopio University Publications, Finlandia, **2007**
17. Todeschini, R.; Consonni, V. Handbook of molecular descriptors. Wiley-VCH, Weinheim, **2000**
18. Cramer, R.D. III; Patterson, D.E.; Bunce, J.D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, 110, págs. 5959-5967
19. Robinson, D.D.; Winn, P.J.; Lyne, P.D.; Richards, W.G. Self-organizing Molecular Field Analysis: a tool for structure-activity studies. *J. Med. Chem.* **1999**, 42, págs. 573-583
20. Goodford, P.J. A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, 28, págs. 849-857
21. Borosy, A.P.; Balogh, B.; Mátyus, P. Alignment free descriptors for quantitative structure-rate constant relationships of [4+2] cycloadditions. *J. Mol. Struct. - Theochem.* **2005**, págs. 169-176
22. Silverman, B.D.; Platt, D.E. Comparative Molecular Moment Analysis (CoMMA): 3D-QSAR without molecular superposition. *J. Med. Chem.* **1996**, 39, págs. 2129-2140



23. Bravi, G.; Gancia, E.; Mascagni, P.; PEGna, M.; Todeschini, R.; Zaliani, A. MS-WHIM, new 3D theoretical descriptors derived from molecular surface properties: a comparative 3D QSAR study in a series of steroids. *J. Comput. Aided Mol. Des.* **1997**, 11, págs. 79-92
24. Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRIND-Independent Descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, 43, págs. 3233-3243
25. Hopfinger, A.J.; Wang, S.; Tokarski, J.S.; Jin, B.; Albuquerque, M.; Madhav, P.J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, 119, págs. 10509-10524
26. Vedani, A.; Briem, H.; Dobler, M.; Dollinger, H.; McMasters, D.R. Multiple conformation and protonation state representation in 4D-QSAR: the Neurokinin-1 receptor system. *J. Med. Chem.* **2000**, 43, págs. 4416-4427
27. Lill, M.A.; Vedani, A. Combining 4D pharmacophore generation and multidimensional QSAR: modeling ligand binding to the Bradykinin B2 receptor. *J. Chem. Inf. Model.* **2006**, 46, págs. 2135-2145
28. Vedani, A.; Dobler, M. 5D-QSAR: The key for simulating induced fit? *J. Med. Chem.* **2002**, 45, págs. 2139-2149
29. Li, Y.; Harte, W.E. A Review of Molecular Modeling Approaches to Pharmacophore Models and Structure-Activity Relationships of Ion Channel Modulators in CNS. *Curr. Pharm. Des.* **2002**, 8, págs. 99-110
30. Kyle, D.J.; Ilyin, V.I. Sodium Channel Blockers. *J. Med. Chem.* **2007**, in press.
31. European Commission's Technical Guidance Document in Support of Commission Directive 93/67/EEC on Risk Assessment for New Notified Substances, Commission Regulation (EC) No. 1488/94 on Risk Assessment for Existing Substances and Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market, Parte III, Capítulo 4, "Use of Quantitative Structure-Activity Relationships (QSAR)", **2003**
32. Dimitrov, S.; Dimitrova, G.; Pavlov, T.; Dimitrova, N.; Patelwicz, G.; Niemela, J.; Mekenyan, O. A stepwise approach for defining the applicability domain of SAR and QSAR models. *J. Chem. Inf. Model.* **2005**, 45, págs. 839-849
33. Hill, T.; Lewicki, P. Statistics Methods and Applications. **2006**, Statsoft, Tulsa, OK.
34. Randić, M. Comparative regression analysis. Regressions based on a single descriptor. *Croat. Chem. Acta.* **1993**, 66, págs. 289-312
35. Randić, M.; Pompe, M. The Variable Connectivity Index $1\chi_f$ versus the Traditional Molecular Descriptors: A Comparative Study of $1\chi_f$ Against Descriptors of CODESSA. *J. Chem. Inf. Model.* **2001**, 41, págs. 631-638



36. Abdi, H. Partial least square (PLS) regression. En *Encyclopedia of social sciences research methods*. Lewis-Beck, M.; Bryman, A.; Liao, T.F. (editores). **2003**, Sage Publications, California, Estados Unidos
37. Schneider, G.; So, S.S. *Adaptative systems in drug design*. Landes Bioscience, Georgetown, Estados Unidos, **2003**
38. Terfloth, L.; Gasteiger, J. Neural networks and genetic algorithms in drug design. *Drug Discov. Today*. **2001**, 6 (Suplemento 2), págs. 102-108
39. Solmajer, T.; Zupan, J. Optimization algorithms and natural computing in drug discovery. *Drug Discov. Today: Technol.* **2004**, 1, págs. 247-252
40. Topliss, J.G.; Edwards, R.P. Chance factors in studies of quantitative structure-activity relationships. *J. Med. Chem.* **1979**, 22, págs. 1238-1244
41. Hawkins, D.M. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, 44, págs. 1-12
42. Yasri, A.; Hartsough, D. Toward and optimal procedure for variable selection and QSAR model building. *J. Chem. Inf. Comput. Sci.* **2001**, 41, págs. 1218-1227
43. Golbraikh, A.; Tropsha, A. Beware of q^2 ! *J. Mol. Graph. Model.* **2002**, 20, págs. 269-276.



Capítulo 3. Descriptores utilizados en la presente tesis.

- Teoría de grafos.
- Matrices y polinomios y espectros derivados del grafo.
Definición de algunos índices topológicos representativos.

Teoría de grafos.

Como ya ha sido presentado, en este trabajo se utilizarán principalmente dos tipos de descriptores como variables independientes de los modelos generados: por un lado, las llamadas “variables indicadoras”; por otro lado, descriptores topológicos derivados de la Teoría de Grafos. Todos ellos pertenecen a la categoría de descriptores que hemos descrito como de baja dimensionalidad, independientes de la orientación y conformación de la estructura química.

Las variables indicadoras indican la presencia/ausencia o frecuencia de un tipo de átomo, grupo funcional o función química. Se trata, por tanto, de descriptores que asumen valores binarios (en el caso que indiquen ausencia o presencia) o en su defecto valores no binarios discretos (cuando indican la frecuencia de aparición en una estructura determinada). Un ejemplo clásico de aplicación de este tipo de variables es el método de Free-Wilson [1]. Esta aproximación supone calcular, mediante análisis de regresión, la contribución a la actividad biológica de cada sustituyente particular en una posición específica, asumiendo que los sustituyentes proveen una contribución constante a la actividad bajo el principio de aditividad. Muchos de los métodos para calcular descriptores fisicoquímicos lo hacen a través de un método por contribución de átomos o grupos; es decir, numerosos modelos QSAR basados en descriptores fisicoquímicos son,



indirectamente, metodologías basadas en variables indicadoras. Tal es el caso del método de Rekker para la estimación del log P [2], según el cual:

$$\log P = \sum_i a_i f_i + \sum \text{términos de interacción}$$

donde a_i es el número de veces que aparece cada átomo o grupo funcional determinado, f_i indica la contribución de éstos a la actividad y los términos de interacción están formados por un número entero de veces una “constante mágica” km . De esta manera, si bien en alguno de los modelos generados se incluirán descriptores fisicoquímicos clásicos, como el log P calculado según Moriguchi o la refractividad molecular de Ghose-Crippen [3,4], los mismos derivan de una combinación de variables indicadoras o índices topológicos, por lo que indirectamente también pueden incluirse dentro del grupo de descriptores de baja dimensionalidad considerados en la presente tesis.

Los descriptores topológicos son descriptores derivados de una rama de las matemáticas conocida como la Teoría de Grafos (como se verá en breve, algunos descriptores topológicos derivan de combinar la Teoría de Grafos con la Teoría de la Información de Shannon). La Teoría de Grafos fue inaugurada en 1736 por Euler, quien resolvió recurriendo a un grafo **el problema de los siete puentes de Königsberg**. Dadas dos islas en el río Pregel, que pasa por la ciudad de Königsberg, habiendo siete puentes que las unen entre sí y a tierra firme, ¿es posible dar un paseo que recorra todos los puentes pasando por cada puente sólo una vez? Euler simplificó el problema asociando a cada puente una arista o eje y a cada uno de los trozos de tierra firme e islas un nodo o vértice (Fig. 4.1). El problema se reducía entonces, simplemente, a observar la conectividad del grafo que se obtiene.

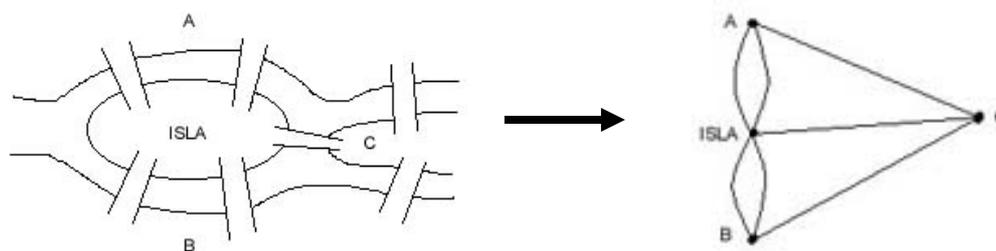


Fig. 4.1. Grafo derivado por Euler para resolver el problema de los siete puentes de Königsberg analizando la conectividad entre las islas y los puntos en tierra firme.

Observando el grafo, Euler concluyó que para poder cumplir los requisitos que exigía el problema por cada vez que se ingresa a un vértice del grafo debe poder salirse del mismo mediante otro eje distinto que aquel por el cual se llegó, con dos excepciones, el vértice por el cual se comienza el recorrido y el vértice por el cual se termina (de los que no es necesario regresar ni salir, respectivamente). En otras palabras, para responder afirmativamente al problema de los puentes: a) todos los vértices excepto dos debían tener un número par de ejes asociados o, b) todos los vértices debían tener un número par de ejes asociados si el vértice de llegada y de salida eran los mismos. Esto no se verifica, ya que la multiplicidad (número de ejes asociados a cada vértice) es en todos los casos tres o cinco, y por lo tanto la respuesta a la pregunta que planteaba el problema es que no es posible, no existe un camino euleriano en el grafo.

La Teoría de Grafos ha encontrado aplicaciones en disciplinas tan diversas como la computación, la lingüística, la criptografía, la ingeniería de los materiales, la biología, las ciencias económicas y sociales y por supuesto, la química, área en el que la forma en que se conectan los átomos en la estructura química puede correlacionarse con diversas propiedades moleculares [5-7].



Los índices topológicos (ITs) son, en el contexto de los estudios QSAR, entidades numéricas derivadas de los grafos asociados a una estructura química que pueden caracterizar aspectos muy diversos de la estructura molecular tales como la forma, el grado de ramificación o ciclación de una estructura química, la distribución de cargas y diversas propiedades fisicoquímicas. En un grafo de una estructura química, **los átomos serán representados como un conjunto de vértices o nodos que denotaremos $V(G)=\{v_1, v_2, \dots, v_n\}$** , siendo n el número de átomos que componen la estructura, mientras que **los enlaces covalentes se representan como un conjunto de ejes o aristas que conectan los vértices $E(G)=\{e_{12}, e_{23}, e_{34}, \dots\}$** . Los átomos de hidrógeno pueden ser considerados explícita o implícitamente. En la Fig. 4.2. se presenta el grafo con hidrógenos implícitos derivado del fenol. Por simplicidad denominaremos a este grafo F para los ejemplos subsiguientes.

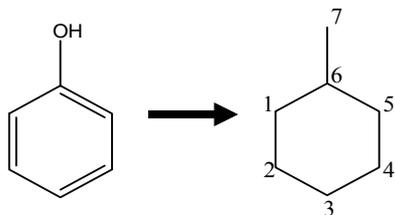


Fig. 4.2. Grafo asociado a la estructura química del fenol. Los ITs son invariantes del grafo y, por lo tanto, es indistinto dónde se inicie la numeración del grafo.

Dos vértices v_i y v_j son adyacentes si existe un eje e_{ij} que los une; se dirá entonces que los vértices v_i y v_j son incidentes del eje e_{ij} . Dos ejes son adyacentes si tienen un vértice común. **El grado del vértice v_i corresponde al número de vértices adyacentes a v_i y será denotado como deg_i** . El conjunto de valores de grados de vértice para un grafo se denota como $\text{Deg}(G)$. Por ejemplo, $\text{Deg}(F)=\{2, 2, 2, 2, 2, 3, 1\}$, considerando la numeración elegida en el grafo de la Fig. 4.2. Un trayecto o *walk* w es una secuencia de vértices y ejes $w=\{v_a, e_{ab}, v_b, e_{bc}, v_c, \dots, v_i, e_{ij}, v_j, \dots, v_n, e_{nm}, v_m\}$ que empieza y termina en



vértices; también puede ser definido como una secuencia de vértices en la que dos vértices que aparecen en forma consecutiva son adyacentes. En un trayecto, cada eje puede ser recorrido más de una vez. Un ejemplo de trayecto en el gráfico del fenol podría ser $\{v_1, e_{12}, v_2, e_{23}, v_3, e_{32}, v_2\}$ o más sencillamente $\{v_1, v_2, v_3, v_2\}$. Un trayecto con en el que el vértice inicial coincide con el final es llamado cerrado o de autorretorno; caso contrario, es llamado abierto o excluyente. **Un camino o *path* es un tipo particular de trayecto en el cual cada vértice aparece sólo una vez.** La longitud de los trayectos o caminos está dada por el número de ejes que los forman. El grafo del fenol, por ejemplo, no tiene caminos de longitud mayor a seis. La excentricidad de un vértice $ecc(v_i)$ se define como la longitud del camino más largo que comienza en v_i ; el diámetro de un grafo será entonces la mayor excentricidad de entre el conjunto de excentricidades de los vértices que componen dicho grafo, mientras que el centro será el vértice de menor excentricidad y el radio será la excentricidad del centro (pudiendo haber, en un mismo grafo, más de un centro).

Matrices, polinomios y espectros derivados del grafo. Definición de algunos índices topológicos representativos.

A fin de utilizar la información contenida en los grafos en relaciones estructura-actividad es necesario “traducir” (al menos parcialmente) el grafo en un valor numérico (descriptor). Para ello lo más común es generar alguna representación algebraica del grafo. Entre estas representaciones encontramos las matrices, polinomios y espectros del grafo. A continuación se describirán, a modo de guía de introducción de la Teoría de Grafos Química, las representaciones algebraicas más comunes del grafo, y algunos de los índices topológicos que pueden obtenerse a partir de las mismas [8]. La



cantidad de representaciones algebraicas e ITs que pueden encontrarse en la literatura es muy extensa y el análisis detallado de cada uno de ellos excede a los alcances del presente trabajo de tesis, por lo que sólo se describirán aquellos usados más extensamente.

La **matriz adyacencia** o **matriz de conectividad** $A=A(G)$ del grafo G con N vértices es la matriz cuadrada simétrica de $N \times N$ cuyos elementos $[A]_{ij}$ se definen como:

$$[A]_{ij} = \begin{cases} 1 & \text{si } i \neq j \text{ y } e_{ij} \in E(G) \\ 0 & \text{si } i = j \text{ ó } e_{ij} \notin E(G) \end{cases}$$

De modo que retomando el ejemplo, la matriz adyacencia del fenol será:

$$A(F) = \begin{vmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{vmatrix}$$

La suma de los elementos de la fila i de la matriz A da el $\text{deg}(v_i)$; del mismo modo, la suma de los elementos de la columna j da el $\text{deg}(v_j)$. Otra forma de interpretar al elemento $[A]_{ij}$ de la matriz adyacencia es pensarlo como el número de trayectos de longitud uno entre los vértices v_i y v_j . Teniendo esto en cuenta, se han definido matrices de k -ésimo poder de la matriz adyacencia con elementos $[A^k]_{ij}$ que representan el número de trayectos de longitud k entre los vértices v_i y v_j . El segundo poder de la matriz adyacencia del fenol $A^2(F)$ está por lo tanto dado por:



$$A^2(F) = \begin{vmatrix} 2 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 2 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 3 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{vmatrix}$$

Siendo los elementos $[A]_{ii}$ de la diagonal de esta matriz el número de trayectos de autorretorno de longitud dos y los elementos fuera de la diagonal el número de trayectos abiertos de longitud dos entre dos vértices i y j . La traza de esta matriz será entonces el **número total de trayectos cerrados de longitud dos que pueden recorrerse en el grafo**, mientras que la suma de los elementos no diagonales constituye el **número de trayectos abiertos de longitud dos totales en el grafo**. La suma de estas dos cantidades es, se desprende, **la suma total de trayectos de longitud dos** (cerrados y abiertos) del grafo. Análogamente pueden obtenerse los mismos descriptores para matrices de orden superior A^k , pudiendo utilizarse éstos y otros descriptores basados en el conteo de trayectos y caminos en estudios QSAR [9-11].

Muchas definiciones de ITs clásicos derivan de la matriz A , entre ellos los índices topológicos M_1 y M_2 de Zagreb y los índices de conectividad molecular χ de Randić:

$$M_1 = \sum_{i=1}^n \text{deg}_i^2$$

$$M_2 = \sum_{\substack{\text{todos} \\ \text{los} \\ \text{ejes}}} \text{deg}_i \text{deg}_j$$



$$\chi = \sum_{\substack{\text{todos} \\ \text{los} \\ \text{ejes}}} (\text{deg}_i \text{deg}_j)^{-1/2}$$

Aunque similar al segundo índice de Zagreb, el índice de conectividad de Randić ha resultado más exitoso en aplicaciones QSAR, y es probablemente el IT más utilizado. Considerando ahora los ejes como caminos de orden uno y generalizando la definición del χ de Randić a caminos de longitud k , Kier, Hall, Randić y Murray obtuvieron una gran variedad de ITs de amplia aplicación [12, 13]:

$${}^k\chi_p = \sum_{\substack{\text{todos} \\ \text{los} \\ \text{caminos}}} (\text{deg}_i \text{deg}_j \dots \text{deg}_k)^{-1/2}$$

Expresiones análogas pueden utilizarse para subgrafos distintos a los caminos (tales como subgrafos cíclicos o ramificados)¹. Dos grafos se definirán como isomórficos si numerando a uno de ellos de alguna manera puede obtenerse la misma matriz de conectividad del segundo. Se dirá que un IT está degenerado si asume el mismo valor para dos grafos no isomórficos; el grado de degeneración de un IT será mayor, por lo tanto, cuantas más estructuras no isomórficas arrojen el mismo valor de ese IT. Los índices de conectividad de órdenes superiores de Kier y Hall presentan un grado de degeneración menor que el de los de los órdenes inferiores.

La segunda matriz que mayor importancia ha tenido en la definición de índices topológicos es **la matriz distancia $D(G)$** . La misma está formada por elementos $[D]_{ij}$ definidos por:

¹ Un subgrafo G' de un grafo G es un grafo cuyos vértices y ejes están contenidos en G .



$$[D]_{ij} = \begin{cases} d_{ij} & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases}$$

donde d_{ij} es el número de ejes que separan los vértices v_i y v_j por el camino más corto. Debajo de estas líneas puede observarse la matriz distancia para el fenol, $D(F)$.

$$D(F) = \begin{vmatrix} 0 & 1 & 2 & 3 & 2 & 1 & 2 \\ 1 & 0 & 1 & 2 & 3 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 0 & 1 & 2 & 3 \\ 2 & 3 & 2 & 1 & 0 & 1 & 2 \\ 1 & 2 & 3 & 2 & 1 & 0 & 1 \\ 2 & 3 & 4 & 3 & 2 & 1 & 0 \end{vmatrix}$$

La semisuma de los elementos de la matriz distancia (o, siendo $D(G)$ por definición simétrica, la suma de los elementos sobre la diagonal principal) da lugar al primer índice topológico conocido, el índice de Wiener (W), que data del año 1947:

$$W = 1/2 \sum_{i=1}^n \sum_{j=1}^n d_{ij}$$

Posteriormente, se han definido una gran cantidad de invariantes del grafo derivadas de la matriz distancia [14-16]. A partir de las matrices A y D Gálvez y colaboradores han definido una nueva matriz topológica M obtenida del producto matricial de A y la matriz formada por la inversa de los cuadrados de los elementos de D [17]. De la matriz M se derivan términos de carga CT definidos como la diferencia entre los elementos $[M]_{ij}$ y $[M]_{ji}$; a su vez, sobre operaciones algebraicas sencillas sobre los CT se derivan los denominados índices de carga de Gálvez.



Otras matrices del grafo menos populares que pueden mencionarse son **la matriz adyacencia de los ejes EA(G)**, definida de manera análoga a la matriz adyacencia A pero considerando la conectividad de los ejes en lugar de la de los vértices [18]; **la matriz laplaciana del grafo**, cuyos elementos diagonales $[L]_{ii}$ son iguales al grado del vértice i y los elementos no diagonales $[L]_{ij}$ iguales a -1 si los vértices v_i y v_j se hallan unidos por un eje o iguales a 0 caso contrario [19]; **la matriz χ** , de elementos $[\chi]_{ij}$ iguales a $(\text{deg}_i \text{deg}_j)$ cuando i y j están unidos por un eje e_{ij} e iguales a 0 caso contrario [20]; **y la matriz detour DD o matriz de distancias topológicas máximas**, de definición idéntica a la matriz distancia (incluso se deriva un índice análogo al índice de Wiener) pero considerando el número de ejes que separan los vértices por el camino más largo en lugar de más corto (nótese que cuando el compuesto considerado es acíclico las matrices D y DD son coincidentes, no así cuando la estructura química del compuesto incluye ciclos) [21].

Hasta aquí se han descrito representaciones algebraicas e ITs derivados del grafo que sólo tienen en cuenta la conectividad entre los átomos, sin considerar la naturaleza de los mismos o de los enlaces covalentes que los unen. Esto implica que, por ejemplo, ninguno de los índices hasta aquí mencionados podría diferenciar al fenol de la anilina o del tolueno, ya que los tres compuestos poseen grafos isomorfos. Esto tiene un motivo histórico: la mayoría de estos índices fueron definidos para caracterizar alcanos y cicloalcanos (los alcanos representan una clase ideal de compuestos para estudiar la variación de propiedades moleculares con la conectividad). No obstante, dado que muchas de las aplicaciones de la Teoría Química de Grafos involucran relaciones QSAR y QSPR entre compuestos que contienen heteroátomos e insaturaciones, se hizo evidente la **necesidad**



de ponderar los vértices y ejes del grafo de distinta manera según las propiedades químicas de los átomos y enlaces asociados a los mismos [22].

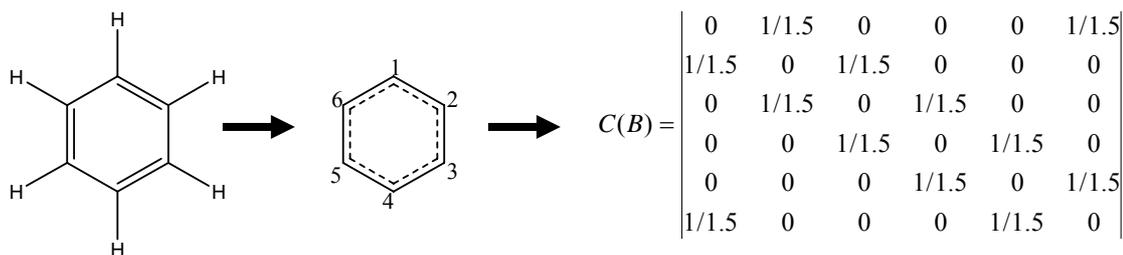
Un grafo con vértices y aristas pesados consistirá entonces no sólo ya en un conjunto de vértices y ejes, sino también un conjunto de símbolos químicos para diferenciar cada tipo de átomo, un set de órdenes topológicos de los ejes B_{ij} , un conjunto de vértices pesados $V_w(w)$ y un conjunto de ejes pesados $E_w(w)$. Usualmente el orden de enlace topológico B_{ij} es igual a: 1 cuando los átomos i y j se encuentran unidos por un enlace simple; 1.5 cuando el enlace es aromático; 2 cuando se trata de un doble enlace y; 3 cuando se trata de un triple enlace.

Se pueden definir entonces nuevas matrices, similares (pero no idénticas) a las ya definidas.

La matriz de conectividad pasa a estar definida de la manera que sigue:

$$[C]_{ij} = \begin{cases} 0 & \text{si } i = j \text{ ó } e_{ij} \notin E(G) \\ 1/B_{ij} & \text{si } i \neq j \text{ y } e_{ij} \in E(G) \end{cases}$$

Según lo dicho, la matriz de conectividad para el benceno (cuyo grafo denotaremos B), está dada por:





Los elementos de las matrices adyacencia y distancia de grafos con vértices y aristas pesados están definidos como:

$$[A(w)]_{ij} = \begin{cases} Vw(w)_i & \text{si } i = j \\ Ew(w)_{ij} & \text{si } i \neq j \text{ y } e_{ij} \in E(G) \\ 0 & \text{si } e_{ij} \notin E(G) \end{cases}$$

$$[D(w)]_{ij} = \begin{cases} Vw(w)_i & \text{si } i = j \\ d(w)_{ij} & \text{si } i \neq j \end{cases}$$

w representa el esquema de pesada de los vértices y ejes. Para obtener $Vw(P)_i$ los vértices pueden ser ponderados según diferentes propiedades atómicas P (por ejemplo, el número atómico Z , la masa atómica A , la electronegatividad relativa X , la polarizabilidad α_v (\AA^3), el radio atómico r (\AA), entre otros. Para ello se aplica el siguiente algoritmo general:

$$Vw(P)_i = 1 - P_c / P_i$$

donde P_c es la propiedad atómica del átomo de C y P_i es la propiedad atómica del átomo i . Los ejes por su parte son ponderados de acuerdo a:

$$Ew(P)_{ij} = P_c P_c / B_{ij} P_i P_j$$

Se introduce el concepto de **valencia del vértice**, que es la suma de las entradas de la matriz adyacencia para cada vértice (cada fila o cada columna) y que no coincide con el orden del vértice, que sigue siendo la cantidad de ejes que inciden en un vértice determinado.



Una forma alternativa de calcular índices topológicos que consideren la naturaleza de los átomos es mediante los índices de conectividad de valencia:

$${}^1\chi^v = \sum_{\substack{\text{todos} \\ \text{los} \\ \text{ejes}}} (\Delta_i^v \Delta_j^v)^{-1/2}$$

que como los índices de conectividad también puede calcularse para órdenes superiores (considerando caminos que incluyan más de un eje). En la expresión anterior:

$$\Delta_j^v = \frac{Z_i^v - H_i}{Z_i - Z_i^v - 1}$$

en donde Z_i es el número atómico del átomo asociado al vértice v_i , Z_i^v representa el número de electrones de valencia del átomo asociado a v_i y H_i el número de átomos de hidrógeno unidos al átomo asociado a v_i .

Como se mencionó, pueden derivarse polinomios y espectros del grafo operando sobre las distintas matrices ya definidas. El polinomio característico del grafo $Ch(G, x)$ es el polinomio característico de la matriz adyacencia derivada de ese grafo:

$$Ch(G, x) = \det(xI - A) = \sum_{n=0}^N c_n x^{N-n}$$

donde I representa la matriz identidad. El determinante de una matriz se calcula mediante:

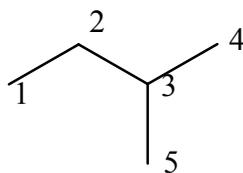
$$\det(A) = \sum_{i=1}^{N!} (-1)^{p_i} P_i A_{11} A_{22} \dots A_{NN}$$



donde P_i es un operador que permuta los índices de las columnas de los elementos de la matriz; la sumatoria comprende $N!$ permutaciones de los índices de columna y p_i indica el número de trasposiciones necesarias para restaurar el orden original. Las raíces del polinomio característico de un grafo corresponderán entonces a los autovalores de la matriz adyacencia [23], y el conjunto de autovalores se denominan espectro del grafo. Nótese que pueden obtenerse polinomios característicos y espectros de cualquier matriz simétrica, por lo que pueden obtenerse polinomios característicos y espectros de todas las matrices del grafo descritas anteriormente en este capítulo. Los autovalores de las distintas matrices han sido utilizados como descriptores topológicos moleculares [8].

Se describirán a continuación dos clases de índices topológicos no derivados de matrices del grafo que se han utilizado en este trabajo: **los descriptores de autocorrelación y los índices de información.**

Considerando que una propiedad de una molécula puede calcularse a partir de contribuciones atómicas (AC_i) mediante un algoritmo de autocorrelación clásico se calculan los productos ($AC_i \times AC_i$, $AC_i \times AC_j$, ...) de la contribución atómica de dos átomos separados por distintas distancias topológicas; la distancia topológica utilizada en cada caso define el orden de la autocorrelación. La suma de los productos de la propiedad atómica considerada para el átomo i por la propiedad atómica de cada uno de los átomos separados por una distancia topológica determinada d nos da la componente d del vector de autocorrelación de la propiedad considerada. Tomemos como ejemplo el grafo correspondiente al isopentano y calculemos el vector de autocorrelación considerando el grado de cada vértice deg_i como propiedad atómica:



donde el deg_1 es igual a 1, el $\text{deg}_2 = 2$, el $\text{deg}_3 = 3$, el $\text{deg}_4 = 1$ y el $\text{deg}_5 = 1$. Las componentes T_i del vector de autocorrelación para el orden de vértice (i denota la distancia topológica) serán:

$$T_0 = 1x1 + 2x2 + 3x3 + 1x1 + 1x1 = 16$$

$$T_1 = 1x2 + 2x3 + 3x1 + 3x1 = 14$$

$$T_2 = 1x3 + 2x1 + 2x1 + 1x1 = 8$$

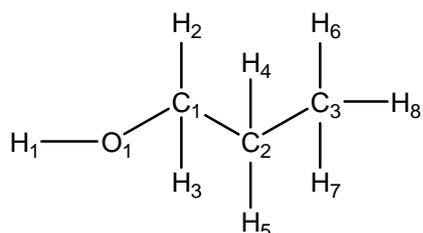
$$T_3 = 1x1 + 1x1 = 2$$

por lo que el vector de autocorrelación será $T=(16,14,8,2)$. Las componentes de orden mayor a tres son nulas por no haber ningún átomo separado de otro por una distancia topológica mayor a tres. Se puede encontrar vectores de autocorrelación para cualquier propiedad de la que se conocen las contribuciones atómicas de los distintos átomos que componen la molécula: volumen de van der Waals, solubilidad en agua, coeficiente de partición octanol-agua, etc. Una de las grandes ventajas de utilizar las componentes de un vector de autocorrelación como descriptores moleculares es la baja degeneración de las mismas [24].

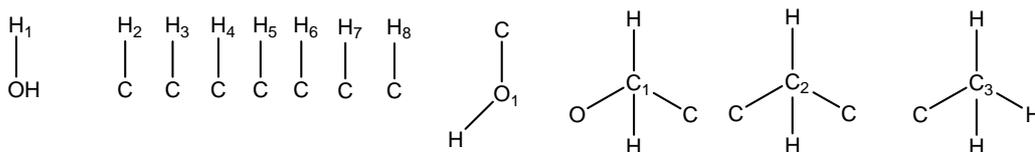
Los índices de información combinan la teoría de grafos con la teoría de la información de Shannon (que sostiene que la información transmitida por una fuente de comunicación o un canal de información es una cantidad mensurable). Los índices de información topológicos cuantifican el grado de redundancia y heterogeneidad en las inmediaciones de algún elemento del grafo (vértices, enlaces),



considerando grafos con hidrógenos explícitos. Los índices de información pueden clasificarse en índices de complejidad topológicos (cuando no se considera la naturaleza de los átomos y enlaces representados en el grafo) y en índices de complejidad química (cuando se considera la naturaleza de los átomos y los enlaces químicos). Considérese como ejemplo el n-propanol:



El conjunto de vecindades o entornos de primer orden está formado por todos los subgrafos conteniendo un vértice determinado y todos los vértices separados de éste por una distancia topológica igual a 1:



Luego, el contenido de información o entropía de información de Shannon puede obtenerse mediante la expresión:

$$IC_k = -\sum_i p_i \log_2 p_i$$

donde p_i es la probabilidad de encontrar al subgrafo de tipo i , extrayendo un elemento al azar del conjunto de subgrafos del orden considerado. En el ejemplo, todos los subgrafos aparecen una única vez en el conjunto de subgrafos, excepto el subgrafo formado por un



carbono y un hidrógeno, que se repite siete veces, por lo cual el contenido de información de orden 1 para el n-propanol está dado por:

$$IC_{1,n\text{-propanol}} = -(1/12 \log_2 1/12 + 7/12 \log_2 7/12 + 1/12 \log_2 1/12 + 1/12 \log_2 1/12 + 1/12 \log_2 1/12)$$

Si la probabilidad de encontrar a cualquiera de los 12 subgrafos fuera idéntica (e igual por lo tanto a 1/12) el contenido de información alcanzaría su valor máximo de $\log_2 12$, y esto significaría que la estructura considerada no tiene subgrafos de orden 1 redundantes (lo cual se asocia a una estructura de mayor “complejidad”). Varios otros índices pueden obtenerse mediante operaciones sencillas sobre el contenido de información [25].

Todos los descriptores utilizados en el transcurso de la presente tesis fueron calculados mediante el uso del programa Dragon [26]. De los más de 1600 descriptores que posee el programa se excluyeron los descriptores tridimensionales (que en rigor de verdad corresponden a lo que en el capítulo anterior se llamó “descriptores 2.5D, independientes de la orientación pero no de la conformación), reteniéndose 877 variables indicadoras y descriptores topológicos más la refractividad molar de Ghose-Crippen, el log P de Moriguchi y el área de la superficie polar calculada en base a fragmentos. Dragón permite calcular 877 descriptores de baja dimensionalidad clasificados según el esquema que se presenta a continuación. Como puede observarse los principios básicos para el cálculo de los descriptores de cada grupo ya han sido abordados a lo largo del presente capítulo:

1. Descriptores constitucionales (peso molecular, número de átomos, número de enlaces múltiples, número de enlaces con libre rotación, entre otros).



2. Descriptores topológicos (entre los que encontramos el índice de Wiener, los índices de Zagreb, la excentricidad y el índice Detour).
3. Conteo de trayectos y caminos.
4. Índices de conectividad (índice de Randic, índices de Kier y Hall, índices de conectividad de valencia, entre otros).
5. Índices de información.
6. Autocorrelaciones 2D.
7. Índices derivados de la Matriz Adyacencia de Ejes (entre ellos autovalores del polinomio característico e índices derivados del espectro de la matriz).
8. Descriptores BCUT (autovalores más altos y más bajos de una matriz A modificada, la matriz Burden, con los elementos de la diagonal ponderados en base a alguna propiedad atómica).
9. Índices topológicos de carga de Gálvez.
10. Índices basados en autovalores de las matrices A y D (esta última pesada de distintas maneras).
11. Conteo de grupos funcionales.
12. Fragmentos centrados en el átomo.

Referencias

1. Free, S.M.; Wilson, J.W. A mathematical contribution to the structure-activity studies. *J. Med. Chem.* **1964**, 1, págs. 395-399
2. Devillers, J. No Free-Lunch Molecular Descriptors in QSAR and QSPR. En Devillers, J.; Balaban, A.T. (editores). En *Topological indices and related descriptors in QSAR and QSPR*. Gordon and Breach Science Publishers, Amsterdam, **1999**, págs. 1-20
3. Moriguchi, I.; Hiroro, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple method of calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull.* **1992**, 40, págs. 127-130
4. Padrón, J.A.; Carrasco, R.; Pellón, R.F. Molecular descriptor based on a molecular refractivity partition using Randic-type graph theoretical invariant.
5. Yen, W.C.K.; Wang, P.P. (editors). *Inform. Sciences* (special issue on graph theory and its applications). **2007**, 177, págs. 2403-2620



6. Keller, A.A. Graph Theory and economic models: from small to large size applications. *Electr. Notes Discrete Math.* **2007**, 28, págs. 469-473
7. Venkata Rao, R.A. A material selection model using graph theory and matrix approaches. *Mat. Sci. Eng. A.* **2006**, 431, págs. 248-255
8. Ivanciuc, O.; Balaban, A.T. The graph description of chemical structures. En Devillers, J.; Balaban, A.T. (editores). En *Topological indices and related descriptors in QSAR and QSPR*. Gordon and Breach Science Publishers, Amsterdam, **1999**, págs. 59-167
9. Rücker, G.; Rücker, C. Walk counts, labyrinthicity, and complexity of acyclic and cyclic graphs and molecules. *J. Chem. Inf. Comput. Sci.* **2000**, 40, págs. 99-106
10. Rücker, G.; Rücker, C. Counts of all walks as atomic and molecular descriptors. *J. Chem. Inf. Comput. Sci.* **1993**, 33, págs. 683-685
11. Rücker, C.; Rücker, G. Mathematical Relation between Extended Connectivity and Eigenvector Coefficients. *J. Chem. Inf. Comput. Sci.* **1994**, 34, págs. 534-538
12. Kier, L.B.; Hall, L.H.; Murray, W.J.; Randić, M. Molecular connectivity. Part 1: Relationship to nonspecific local anesthesia. *J. Pharm. Sci.* **1975**, 64, págs. 1971-1974
13. Kier, L.B.; Murray, W.J.; Randić, M.; Hall, L.H. Molecular connectivity. Part 5: Connectivity series concept applied to density. *J. Pharm. Sci.* **1976**, 65, págs. 1806-1809
14. Balaban, A.T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, 89, págs. 399-404
15. Balaban, A.T.; Balaban, T.S. New vertex invariants and topological indices of chemical graphs based on information on distances. *J. Math. Chem.* **1991**, 11, págs. 79-105
16. Estrada, E.; Ivanciuc, O.; Gutman, I.; Gutiérrez, A.; Rodríguez, L. Extended Wiener Index. A new set of descriptors for quantitative structure-property studies. *New J. Chem.* **1998**, 22, págs. 819-822
17. Gálvez, J.; García, R.; Salabert, M.T.; Soler, R. Charge Indexes. New topological descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, 34, págs. 520-525
18. Estrada, E.; Ramírez, A. Edge adjacency relationships in molecular graphs and molecular topographic descriptors. Definition and QSAR applications. *J. Chem. Inf. Comput. Sci.* **1996**, 36, págs. 837-843
19. Trinajstić, N.; Babić, D.; Nikolić, C.; Plavšić, D. The laplacian matrix in chemistry. *J. Chem. Inf. Comput. Sci.* **1994**, 34, págs. 368-376
20. Randić, M. Similarity based on extended basis descriptors. *J. Chem. Inf. Comput. Sci.* **1992**, 32, págs. 686-696
21. Randić, M. On characterization of cyclic structures. *J. Chem. Inf. Comput. Sci.* **1997**, 37, págs. 1063-1071
22. Ivanciuc, O.; Ivanciuc, T.; Balaban, T.A. Vertex- and edge-weighted molecular graphs and derived structural descriptors. En *Topological*



- indices and related descriptors in QSAR and QSPR*. Gordon and Breach Science Publishers, Amsterdam, **1999**, págs. 169-220
23. Szabo, A.; Ostlund, N.S. *Modern Quantum Chemistry. Introduction to Advanced Electronic Structure Theory*. Chapter 1. Mathematical Review. Dover Publications, New York, **1996**, págs. 2-38
24. Devillers, J. Autocorrelation descriptors for modeling (eco)toxicological endpoints. En *Topological indices and related descriptors in QSAR and QSPR*. Gordon and Breach Science Publishers, Amsterdam, **1999**, págs. 595-612
25. Basak, S.C. Information theoretic indices of neighborhood complexity and their applications. En *Topological indices and related descriptors in QSAR and QSPR*. Gordon and Breach Science Publishers, Amsterdam, **1999**, págs. 563-593
26. Dragon Academic versión 4.0, software para el cálculo de descriptores moleculares. Milano Chemometrics, Milán, **2003**



Capítulo 4. Screening Virtual.

- ✦ Definición. ¿Por qué utilizar Screening Virtual?
- ✦ Clasificación.
- ✦ Perspectivas.
- ✦ Aproximaciones utilizadas en el presente trabajo.

Definición. ¿Por qué utilizar SV?

Se denomina SV al conjunto de métodos computacionales destinados a seleccionar, de entre grandes bases de datos o colecciones virtuales de estructuras químicas, aquellos compuestos que por reunir ciertas características estructurales son candidatos a presentar una propiedad o actividad biológica determinada.

El universo de estructuras químicas sintéticamente accesibles en la actualidad se ha estimado en 10^{100} [1]. La capacidad aparentemente ilimitada de la Química moderna para sintetizar nuevos compuestos se refleja en el crecimiento exponencial del número de compuestos registrados en el Chemical Abstract Service en las últimas cuatro décadas. En 1965, la cantidad de compuestos registrados ascendía a 212.000 sustancias. Treinta años después la cifra ascendía a 16 millones, con un millón de estructuras registradas cada año (a razón de dos nuevos compuestos descubiertos por minuto); esta evolución de la capacidad sintética se ha acentuado aún más desde entonces con el desarrollo de la combinatoria química [2]. El aspecto positivo de esta situación es que es altamente probable encontrar, en un universo químico de abundancia y diversidad impensable, compuestos químicos con las bioactividades que la Medicina demanda. Se comprenderá, sin embargo, que el muestreo sistemático de todas las estructuras conocidas en búsqueda del fármaco ideal dentro de cada categoría



terapéutica es inviable por cuestiones de tiempo y recursos disponibles. Frente a este desafío la creatividad del hombre ideó nuevas estrategias para explorar y aprovechar la inmensa, y creciente, diversidad química, tales como el SV y el HTS. Como se describió en la introducción, las estrategias actuales de la Química Medicinal llaman a una búsqueda racional y (en parte debido a que la Química Medicinal lidia con necesidades sanitarias urgentes de la población, por un lado, y al paradigma de eficiencia de la época, por otro) veloz de nuevos agentes terapéuticos.

El objetivo fundamental del SV es por tanto identificar, en un tiempo relativamente breve, nuevas estructuras líderes que permitan responder adecuadamente a las necesidades de la Medicina. Estas necesidades pueden ser de muy diversa naturaleza:

- Necesidad de estructuras que interaccionen con nuevos blancos terapéuticos (por ejemplo, un receptor o enzima cuya existencia, función o estructura se hayan dilucidado recientemente).
- Necesidad de agentes terapéuticos más potentes (por ejemplo, que presenten mayor biodisponibilidad o con mayor capacidad de interacción con el blanco molecular) o menos tóxicos (por presentar mayor selectividad respecto al blanco molecular) que los ya conocidos.
- Necesidad de agentes terapéuticos más accesibles en términos económicos (el precio de los medicamentos, no obstante, obedece también, claramente, a cuestiones políticas y económicas que superan los alcances de esta tesis y, en general, del químico medicinal).



El hecho de poder estudiar grandes colecciones de compuestos, de entre cientos de miles y millones de estructuras, hace presuponer (como se demostrará en la parte experimental de la tesis) que **pueden encontrarse nuevos agentes terapéuticos estructuralmente diversos, lo que encierra una alta probabilidad de poder responder simultáneamente, con una sólo estructura química, a todas o buena parte de las necesidades descritas.** El SV posee varias ventajas que vale la pena mencionar:

- **Minimización de los costos inherentes al desarrollo de un nuevo fármaco.** Las diferentes etapas que debe superar un fármaco durante el proceso de desarrollo previamente a su uso en humanos insumen entre 500 y 2000 millones de dólares y un tiempo promedio de 12 años [3,4]. Esas cifras incluyen no sólo la inversión real, sino también el costo de capital, esto es, los intereses que podrían devengarse colocando, durante el tiempo de desarrollo, el monto total invertido en una inversión segura (alternativamente, puede entenderse como los intereses que pagará el laboratorio productor si obtiene el dinero necesario para el desarrollo mediante un préstamo). La mayor parte del capital invertido se gasta en las fases de desarrollo preclínico y, sobre todo, clínico, es decir, en los ensayos en animales y humanos que deben llevarse a cabo para garantizar que el nuevo fármaco es eficaz y seguro. **Mediante el SV pueden acortarse los tiempos de desarrollo y maximizarse las probabilidades de éxito en las etapas preclínica y clínica, optimizando el uso de recursos (materiales y humanos) y del recurso tiempo.** Conviene subrayar que el SV no pretende reemplazar a los ensayos in vivo e in vitro, sino minimizar las probabilidades de fracaso en esas instancias. **Los ensayos en silicio y los ensayos in vitro e in vivo deben entenderse como herramientas complementarias y la**



validación definitiva del modelo teórico será siempre la comparación con el resultado experimental. La velocidad del SV y la posibilidad de elegir, entre los candidatos seleccionados, compuestos con perfiles fármaco y toxicológico conocidos permiten responder a las necesidades urgentes de la población afectada por una patología determinada.

- Adicionalmente, el SV tiene un fuerte fundamento bioético: las pautas éticas internacionales para estudios que involucran animales y humanos coinciden en que, si bien unos y otros son inevitables, **debe priorizarse el uso de modelos matemáticos y computacionales siempre que sea posible** y reducir el número de animales y humanos participantes al mínimo necesario desde el punto de vista estadístico para obtener resultados válidos [5,6].

Clasificación.

Las distintas técnicas de SV se diferencian, fundamentalmente, en los requisitos de cada una de ellas para llevar a cabo una búsqueda significativa [7-9]. En principio, las técnicas de SV **se clasifican en dos grandes grupos según se enfoquen en el receptor o en el ligando**. Usando el paradigma clásico de la llave y la cerradura, el SV basado en el receptor consiste en investigar la forma del ojo de la “cerradura” (las características estructurales del blanco molecular, sea un receptor, una enzima o un blanco molecular de otra naturaleza) de manera de encontrar nuevas “llaves que la abran”. El SV basado en el ligando implica investigar llaves que, se sabe, abren la cerradura de interés (ligandos naturales o fármacos ya conocidos que tienen el blanco molecular de interés), para detectar el patrón común en las mismas e identificar llaves nuevas.



El SV basado en el receptor (docking) requiere información previa de la estructura tridimensional del blanco molecular [10,11]. Esta información no siempre está disponible; en nuestro caso, ya se ha visto que el blanco molecular que nos interesa (los canales de sodio voltaje-operados del SNC) es un receptor sumamente complejo con múltiples dominios transmembrana y múltiples sitios de unión a fármacos. Es, por lo tanto, un receptor particularmente difícil de modelar. Por otro lado, no se han desarrollado todavía técnicas de docking enteramente automáticas, por lo cual no es la primera opción cuando se desea realizar un SV de grandes bases de datos de estructuras químicas, como la que hemos abordado en el presente trabajo.

Dentro de las técnicas basadas en el ligando podemos mencionar tres grandes grupos: **las técnicas basadas en superposición de ligandos, las técnicas basadas en descriptores y las técnicas basadas en similitud.**

Las **técnicas de SV basadas en superposición** consisten justamente en superponer cada una de las estructuras de la base de datos con una estructura de referencia (un ligando o, si se ha identificado, un farmacóforo). La superposición dispone sectores de la estructura de la base de datos sobre sectores químicamente similares de la estructura de referencia, preferentemente alineando grupos involucrados en interacciones de corto alcance (por ej. puentes de H). Otra alternativa para definir la orientación de la estructura de la base de datos sobre la estructura de referencia es realizar un mapeo de algún campo de fuerza (por ejemplo, el campo electrostático) y luego proceder a la superposición de las dos estructuras consideradas. Tanto la molécula de referencia como la molécula de la base de datos podrán considerarse como entidades rígidas o flexibles. Las limitaciones del



SV basado en superposición son, se desprende, similares a las que ya se describieron para el 3D-QSAR: a) **necesidad de definir conformaciones probables de las moléculas consideradas** (e idealmente, **necesidad de identificar la conformación activa** y la conformación de mínima energía de la molécula de la base de datos, definir si es posible que la molécula de la base de datos asuma la conformación activa y, en tal caso, calcular la probabilidad de ese evento como la diferencia de energía entre la conformación de mínima y la conformación activa); b) **necesidad de definir algún criterio de orientación de las estructuras**. El pre-procesamiento de las estructuras y la superposición en sí son procesos computacionalmente costosos. Si bien podría argumentarse que el costo computacional es un factor relativo considerando el ininterrumpido desarrollo de procesadores cada vez más potentes, debe recordarse que la cantidad de estructuras químicas crece exponencialmente, mucho más rápido de lo que crece la velocidad de cálculo; a modo de ejemplo, considérese que las bases de datos de libre acceso de estructuras químicas que se utilizaron en la presente tesis incorporan anualmente alrededor de un millón de nuevas estructuras; actualmente se considera que un método SV debe ser capaz de procesar 10 millones de estructuras en semanas [12]. Algunos ejemplos de metodologías de SV por superposición son FlexS [13], GASP [14] y MIMIC [15].

El **SV basado en descriptores** es, comparativamente, una técnica mucho más eficiente. Esta metodología se basa en predecir la categoría o el valor de la propiedad biológica de interés de las moléculas de la base de datos mediante una ecuación QSAR. Las bases de la aproximación QSAR y la definición de un gran número de descriptores, así como las ventajas y desventajas de cada familia de descriptores, se han discutido ya con cierto detalle y no se volverá sobre ello. El único requisito de esta metodología es que debe



conocerse un número de ligandos lo suficientemente grande como para derivar un modelo QSAR; estos ligandos, dado que el modelo se aplicará en la predicción de la actividad de estructuras diversas, deben ser estructuralmente diversos.

El **SV basado en similitud** consiste en la comparación de “huellas dactilares 2D” (2D fingerprints) de las moléculas de la base de datos con una molécula de referencia. Las huellas dactilares son una secuencia binaria que codifica la presencia o ausencia de una serie de fragmentos o subestructuras (por ejemplo, pares, triadas y otras secuencias de átomos) [16]. Para cuantificar la similitud se recurre a los llamados coeficientes de similitud, entre los que encontramos el coeficiente de Tanimoto ($S = \frac{c}{a+b-c}$), el de Cosine ($S = \frac{c}{\sqrt{ab}}$) y el de Hamming ($S = a+b-2c$) (a representa el número de subestructuras presentes en el compuesto A; b el número de subestructuras presentes en el compuesto B y c el número de subestructuras presentes simultáneamente en ambos (intersección de los conjuntos de subestructuras de A y B)). El coeficiente de Tanimoto es claramente el más utilizado en búsquedas por similitud; la definición del mismo presentada más arriba corresponde a la forma binaria del dicho coeficiente, pero otras formas han sido definidas [17]. La gran ventaja del SV basado en similitud es que la búsqueda puede realizarse utilizando una única molécula de referencia, por lo que es de suma utilidad cuando no se conoce un número de ligandos suficiente para desarrollar un modelo QSAR. Adicionalmente, casi no se requieren decisiones por parte del operador humano (en contraste con lo que ocurre con la superposición o en las etapas de generación de un modelo QSAR) y es el método menos costoso, computacionalmente hablando. En la Tabla 5.1 se presenta un resumen de las aproximaciones descritas y los requisitos de cada metodología.



Tabla 5.1. Comparación de los requisitos de las distintas metodologías de SV.

Tipo de Screening Virtual		Requisitos
Basado en el receptor		<ul style="list-style-type: none">• Estructura 3D del blanco molecular• Sitio activo identificado y preparado (adición de hidrógenos, determinación de los estados de protonación de los residuos, etc).
Basado en el ligando	Basado en superposición	<ul style="list-style-type: none">• Conformación probable de las moléculas a superponer• Criterio de orientación de las moléculas de prueba sobre la molécula de referencia• Idealmente, identificación de la conformación activa para calcular la diferencia de energía entre la conformación probable de las moléculas que se superponen y la conformación activa
	Basado en descriptores	<ul style="list-style-type: none">• Número mínimo de ligandos conocidos y estructuralmente diversos para derivar un modelo QSAR significativo
	Basado en similitud	<ul style="list-style-type: none">• Un ligando conocido

Pese a que tradicionalmente se considera al SV basado en el receptor como la alternativa de primera elección cuando se dispone de la estructura del blanco molecular a resoluciones atómicas altas, recientes estudios sistemáticos han reivindicado al SV basado en el ligando [8,12,18]. En particular, el trabajo de Zhang y Muegge (comparando los resultados del SV mediante docking con técnicas de SV basadas en similitud 3D y 2D para siete blancos moleculares conocidos, entre ellos la COX-2, el receptor de estrógeno y una



proteasa del virus HIV), indican que en la mayoría de los casos el desempeño del screening basado en el ligando supera al del screening basado en docking.

Perspectivas.

El proceso de desarrollo de un fármaco es como se describió un proceso largo y costoso, y está caracterizado por una alta tasa de atrición (un 90% de atrición hacia el final de las fases clínicas de evaluación; sólo uno de cada diez proyectos de desarrollo de fármacos culmina en un nuevo medicamento en el mercado). Una buena parte de esas atriciones (alrededor de un 60%) se debe a problemas vinculados con la Administración, Distribución, Metabolización y Eliminación (ADME) y a dificultades relacionadas con la toxicidad de los agentes terapéuticos en desarrollo; las atriciones asociadas a problemas de eficacia, mientras tanto, representan no más del 30% [19]. En la Fig. 5.1. se presentan las razones de fracaso de proyectos de desarrollo de nuevos fármacos; los valores han sido tomados de Schuster y col. (2005).

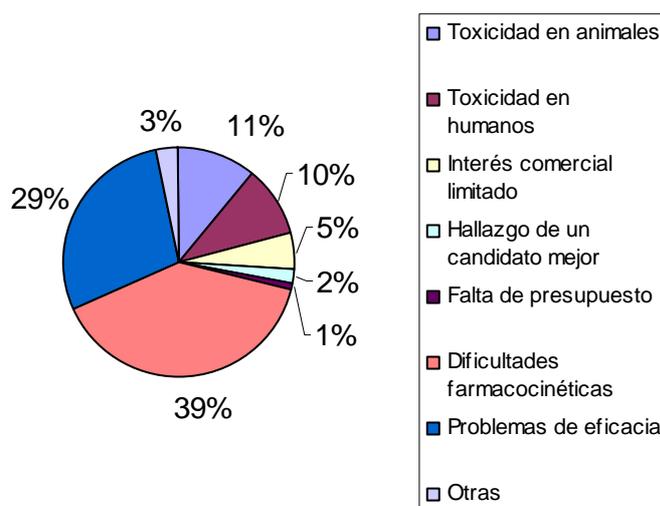


Fig. 5.1. Motivos de terminación prematura de proyectos de desarrollo de nuevos fármacos entre 1964 y 1985.



Esta situación ha motivado un cambio en el paradigma de desarrollo de nuevos fármacos en la última década. El paradigma anterior concentraba su atención en el descubrimiento de compuestos potentes, esto es, en la eficacia del nuevo agente terapéutico, concentrándose en las características ADME/Tox posteriormente, cuando ya se había invertido un capital considerable en el proyecto. El paradigma actual propone incluir filtros ADME/Tox en las primeras etapas del desarrollo, es decir, dirigir el proceso de screening inicial no sólo a la identificación de candidatos activos, sino también con un perfil farmacocinético y toxicológico adecuado [20,21]. En este trabajo se ha procurado generar y aplicar, siguiendo la tendencia actual, filtros computacionales que retengan no sólo estructuras con actividad sino además seguras y biodisponibles a nivel central luego de su administración oral.

Por último, cabe destacar dentro de las nuevas estrategias de SV el **uso de metodologías por consenso** tanto en SV basado en el ligando como en SV basado en el receptor. Las metodologías por consenso seleccionan estructuras identificadas como activas no por uno sino por varios modelos QSAR independientes, sacrificando estructuras correctamente clasificadas como activas por alguno de los modelos con el objeto de reducir el número de falsos positivos (Fig. 5.2) [12,22,23]. Otra tendencia es la **aplicación de metodologías combinadas** que emplean en las etapas iniciales del proceso de SV los filtros más eficientes desde el punto de vista computacional (por ejemplo, SV basado en similaridad o en descriptores 2D), reservando los filtros más costosos para las etapas finales (SV por superposición o SV basado en el receptor). Se consigue de esta manera una especie de sinergismo entre las distintas metodologías aplicadas, mediante metodologías que combinan los mejores aspectos de cada una de las aplicaciones mencionadas.

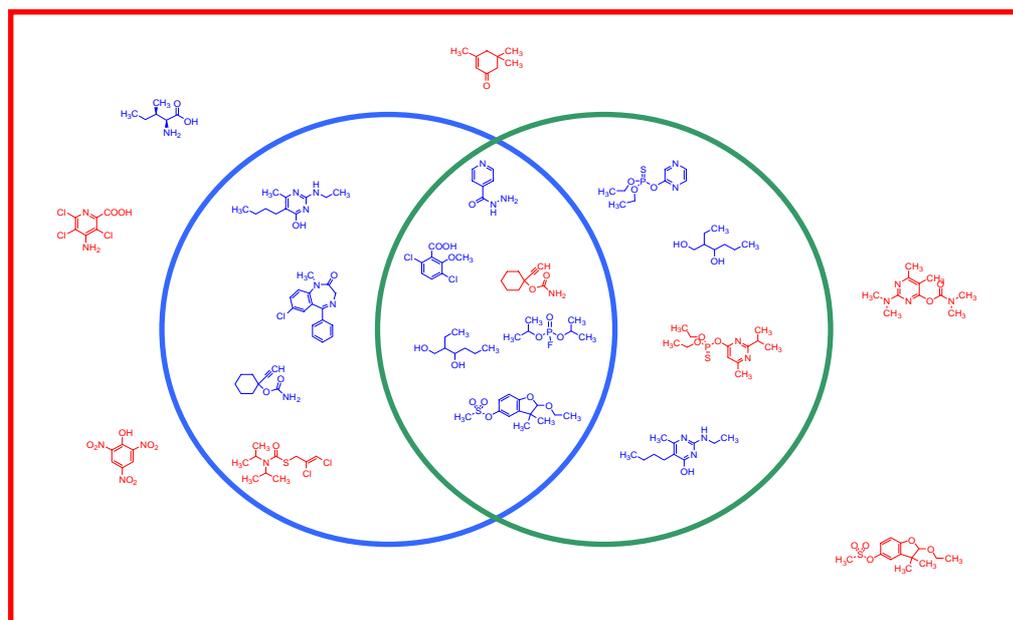


Fig. 5.2. Esquema de los fundamentos de las metodologías QSAR por consenso. Las estructuras en azul representan compuestos con la actividad que se busca; las estructuras en rojo compuestos inactivos. El conjunto en rojo representa el conjunto total de estructuras analizadas mediante SV. Los conjuntos azul y verde representan las estructuras señaladas como activas por el modelo QSAR 1 y el modelo 2, en ese orden (siendo que el modelo 1 y el modelo 2 modelan la actividad buscada y han sido generados de manera independiente). LA intersección de ambos modelos contiene las estructuras seleccionadas mediante QSAR por consenso. El modelo 1, en el ejemplo, incluye un 20% (2/10) de falsos positivos; el modelo 2 incluye otro 20% (2/10) de falsos positivos, pero la intersección incluye sólo un 16.7% (1/6). La intersección se ha enriquecido por lo tanto en compuestos activos si se compara con cada conjunto considerado de manera separada. Este enriquecimiento sacrifica compuestos activos que nos encontramos en la intersección. El conjunto universal de las estructuras sometidas a screening incluye, en el ejemplo, un falso negativo.

Aproximaciones utilizadas en el presente trabajo.

La estrategia utilizada en el presente trabajo de tesis se basa fundamentalmente en el uso de SV basado en descriptores de baja dimensionalidad, con aplicación posterior de filtros ADME/Tox para retener estructuras biodisponibles a nivel central tras administración oral y seguras. Sin embargo, en los anexos I y II se presentan aplicaciones de modelos basados en coeficientes de similitud y técnicas combinadas 2D-3D QSAR. Se han explorado, de esta manera, los tres tipos de SV basado en el ligando que se describieron en este



capítulo, utilizando además las estrategias más modernas de SV que contemplan la asociación sinérgica de diferentes técnicas.

Adicionalmente, el anexo IV presenta una aplicación original del SV: la búsqueda en grandes bases de datos, mediante SV, no ya de estructuras bioactivas sino de reactivos químicos específicos.

Referencias

1. Schneider, G.; So, S.S. *Adaptative systems in drug design*. Chapter 1. A conceptual framework. **2003**, Landes Bioscience, Georgetown, Estados Unidos
2. Freemanthe, M. What are the limits of Chemistry? *Chem. Eng. News*. **1998**, 27, pág. 45
3. Dimasi, J.A.; Hansen, R.W.; Grabowsky, H.G. The price of innovation: new estimates of drug development cost. *J. Health Econ.* **2003**, 22, págs. 151-185
4. Adams, C.P.; Branster, V.V. Estimating The Cost Of New Drug Development: Is It Really \$802 Million? *Health Affair*. **2006**, 25, págs. 420-428
5. Council for International Organizations of Medical Sciences en colaboración con la OMS. CIOMS international ethical guidelines for biomedical research involving human subjects. **2002**
6. Council for International Organizations of Medical Sciences. International guiding principles for biomedical research involving animals. **1985**
7. Lenagauer, T.; Lemmen, C.; Rarey, M.; Zimmermann, M. Novel technologies for virtual screening. *Drug Discov Today*. **2004**, 9, págs. 27-34
8. Sheridan, R.P.; Kearsley, S.K. Why do we need so many chemical similarity search methods? *Drug Discov. Today*. **2002**, 7, págs. 903-911
9. Jalaie, M.; Shanmugasundaram, V. Virtual screening: are we there yet? *Mini Rev. Med. Chem.* **2006**, 6, págs. 1159-1167
10. Sousa, S.F.; Fernandes, P.A.; Ramos, M.J. Protein-ligand docking: current status and future challenges. *Proteins*. **2006**, 65, págs. 15-26
11. Lyne, P.D. Structure-based virtual screening: an overview. *Drug Discov. Today*. **2002**, 7, págs. 1047-1055
12. Muegge, I.; Oloff, S. Advances in virtual screening. *Drug Discov. Today: Technol.* **2006**, 3, págs. 405-411



13. Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: a method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, 41, págs. 4502-4520
14. Jones, G.; Willett, P.; Gwen, R.C. A genetic algorithm for flexible. molecular overlay and pharmacophore elucidation. *J. Comput. Aided Mol. Des.* **1995**, 9, págs. 532-549
15. Krongrad, A.; Clark, L.C.; Hong, L.; Soloway, M.S.; Shenghan, L.; Mestres, J.; Rohrer, D.C.; Maggiora, G.M. A molecular field-based similarity approach to pharmacophoric pattern recognition. *J. Mol. Graph. Model.* **1997**, 15, págs. 114-121
16. Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov. Today.* **2006**, 11, págs. 1046-1053
17. Chen, X.; Reynolds, C.H. Performance of similarity measures in 2D fragment-based similarity searching: comparison of structural descriptors and similarity coefficients. *J. Chem. Inf. Comput. Sci.* **2002**, 42, págs. 1407-1414c
18. 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, págs. 1536-1448
19. Schuster, D.; Laggner, C.; Langer, T. Why drugs fail - a study on side effects in new chemical entities. *Curr. Pharm. Design.* **2005**, 11, págs. 3545-3559
20. Yu, H.; Adedoyin, A. ADME/Tox in drug discovery. Integration of experimental and computational technologies. *Drug Discov. Today.* **2003**, 8, págs. 853-861
21. Modi, S. Positioning ADME in silico tools in drug discovery. *Drug Discov. Today.* **2004**, 9, págs. 14-15
22. Baurin, N.; Mozziconacci, J.C.; Arnoult, E.; Chavatte, P.; Marot, C.; Morin-Allory, L. 2D QSAR consensus prediction for high-throughput virtual screening. An application to COX-2 inhibition modeling and screening of the NCI database. *J. Chem. Inf. Comput. Sci.* **2004**, 44, págs. 276-285
23. Charifson, P.S.; Corkery, J.J.; Murcko, M.A.; Walters, P. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, 42, págs. 5100-5109

PARTE II

PARTE EXPERIMENTAL



Capítulo 5. Función discriminante.

- Fundamentos del uso de la función discriminante.
- Pasos del modelado. Aplicación en Screening Virtual.

Fundamentos del uso de la función discriminante.

En este trabajo de tesis se intentó, originalmente, derivar un modelo QSAR de la forma $\log ED_{50} = a_0 + a_1 d_1 + \dots + a_m d_m$ para la predicción de actividad anticonvulsiva en el ensayo MES (asociado a bloqueo de canales de sodio voltaje-operados), siendo $\{d_1, d_2, \dots, d_m\}$ un conjunto de descriptores de baja dimensionalidad, constitucionales y/o topológicos. Sin embargo, como ya se discutió en el capítulo 2 de la tesis, el desarrollo de un buen modelo QSAR con una variable dependiente cuantitativa exige a) la existencia de un grupo de compuestos, en lo posible estructuralmente diversos, para los cuales el valor de la variable dependiente haya sido determinado experimentalmente y b) que los valores experimentales y observados de la variable dependiente se hallen bien distribuidos a lo largo de tres o cuatro órdenes logarítmicos.

Esta última condición supuso una limitación importante al momento de desarrollar el modelo QSAR para la identificación de candidatos anti-MES: **casi no se han reportado en literatura valores de ED_{50} para compuestos poco activos.** Esta ausencia de datos experimentales de ED_{50} para compuestos poco activos obedece fundamentalmente a dos cuestiones:



1. En general, el sistema científico parece valorar más los datos positivos que los datos negativos, sobre todo en el área de desarrollo de nuevos fármacos. Si un fármaco no se demuestra particularmente promisorio en la primera etapa de los programas internacionales para el desarrollo de nuevos agentes anticonvulsivos, el investigador en general abandona la evaluación de ese fármaco sin determinar la ED_{50} ni otros datos farmacológicos o farmacocinéticos.
2. Existen limitaciones fisicoquímicas para el ensayo en animales de dosis elevadas de un fármaco. El número de solventes que pueden utilizarse como vehículo para la administración de un fármaco a animales es limitado, ya que el vehículo utilizado deber ser, idealmente, inocuo e inerte desde el punto de vista farmacológico. En el caso del ensayo de fármacos que actúan a nivel del SNC el número de solventes adecuados es, por los requisitos anteriores, aún más reducido que para fármacos de acción periférica, ya que los alcoholes tienen, según la dosis, efectos excitatorios o depresores sobre el SNC. Si se usara un alcohol como vehículo, sería imposible saber si el efecto anticonvulsivo se debe a la droga administrada o a los efectos sedantes del vehículo. La capacidad para ensayar altas dosis de un fármaco anticonvulsivo estará limitada entonces por la solubilidad de dicho fármaco en los muy escasos vehículos hidrofílicos y farmacológicamente inertes disponibles. Aún cuando se admitiera la aplicación de suspensiones, la absorción de fármacos en suspensión es un proceso complejo y errático que condicionaría la reproducibilidad del dato obtenido.

Para superar esta limitación de escasez de datos experimentales disponibles en la zona del espacio químico correspondiente a los compuestos poco activos se recurrió a una técnica conocida como



Análisis Lineal Discriminante (ALD). El ALD es una metodología sencilla ampliamente utilizada en el desarrollo de nuevos fármacos de las más diversas categorías terapéuticas [1-5]. Consiste en derivar, mediante el método de los cuadrados mínimos, una función discriminante (df) capaz de distinguir entre dos o más clases de compuestos (por ejemplo, compuestos activos e inactivos) de la forma:

$$df = a_0 + a_1d_1 + a_2d_2 + \dots + a_nd_n$$

donde una vez más $\{d_1, d_2, \dots, d_n\}$ se refiere a un conjunto de descriptores capaces de discriminar las clases o categorías propuestas y $\{a_1, a_2, \dots, a_n\}$ es el conjunto de coeficientes de regresión; como se ve, el ALD es muy similar a la regresión lineal múltiple excepto por el hecho de que df , la variable dependiente, es ahora una variable categórica que puede asumir tantos valores como categorías se estén considerando; usualmente, en el caso de dos categorías, asumirá valores de -1 ó 0 para la categoría inactiva y 1 para la categoría activa.

La gran ventaja del ALD es, en nuestro caso, que la categoría inactiva se compondrá por un conjunto de compuestos bioactivos pertenecientes a otras categorías distintas a la de interés (es decir, fármacos sin reporte de actividad anticonvulsiva en el ensayo MES), superando las limitaciones de las que hablábamos en párrafos anteriores. Desde luego, la hipótesis fundamental de un ALD de estas características es que las drogas incluidas en la categoría inactiva no poseen actividad. Usualmente esta hipótesis no se corrobora (esto es, no se ensayan los fármacos incluidos en la categoría inactiva para



verificar fehacientemente que no tienen la actividad de interés¹) y la incertidumbre respecto a que efectivamente dichos compuestos sean inactivos se constituye en la principal fuente de error del método. Desde luego, garantizando a) un número de compuestos “inactivos” adecuado en el conjunto de entrenamiento; b) una gran diversidad estructural en el conjunto de entrenamiento y; c) una procedimiento de validación adecuado, la probabilidad de errores debido a esa hipótesis incierta puede minimizarse a valores aceptables.

Pasos del modelado. Aplicación en SV.

Se describirá a continuación, en detalle, la generación y validación del modelo QSAR utilizado principalmente en este trabajo, siguiendo el orden y las pautas descritas en el capítulo 2 (apartado “Etapas generales de un estudio QSPR/QSAR”).

1. **Generación del conjunto de entrenamiento.** Se generó un conjunto de entrenamiento estructuralmente diverso, formado por un conjunto de 21 compuestos que integran la categoría ACTIVA del conjunto de entrenamiento y 28 compuestos que componen la categoría INACTIVA. Ambas categorías, en ese orden, pueden observarse en las figs. 6.1 y 6.2. Los 21 compuestos de la categoría ACTIVA son fármacos anticonvulsivantes activos en el ensayo MES realizado en ratones, cuya ED_{50} es menor a 200 mg/kg (administración intraperitoneal), de uso clínico o en etapa clínica de evaluación a la fecha de generar la *df*. Se utilizó el criterio de incluir en el conjunto de entrenamiento sólo fármacos anticonvulsivos de uso clínico o en fase clínica de evaluación para, indirectamente, “capturar” en el modelo QSAR generados características estructurales no sólo

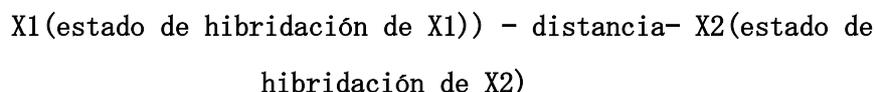
¹ Si se medita al respecto, evaluar un gran número de fármacos para corroborar su inactividad iría en contra del principio de optimización de ensayos biológicos que es la razón de ser del screening virtual.



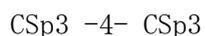
asociadas a la interacción con el blanco molecular sino también a otras cuestiones que hacen a la efectividad y seguridad de un fármaco, tales como la biodisponibilidad a nivel central y la ausencia de efectos adversos graves que pongan en peligro la vida del paciente. La categoría INACTIVA está formada por 27 fármacos pertenecientes a otras categoría terapéuticas, entre las que podemos mencionar: antiinflamatorios no esteroideos, analgésicos, antiulcerosos, antihipertensivos, antibióticos, antivirales, antiparasitarios, antitusivos y broncodilatadores. La Tabla 6.1 presenta los nombres de los compuestos que integran el conjunto de entrenamiento; la categoría terapéutica de los fármacos que integran la categoría INACTIVA ha sido extraído de la 13ra edición del Merck Index [6], o en su defecto de Pubchem [7]. A fin de garantizar la diversidad estructural del conjunto de entrenamiento, y en particular de la categoría inactiva, se siguieron los siguientes criterios: a) se incluyeron en la categoría INACTIVA fármacos de diversas categorías terapéuticas. El hecho de que los fármacos pertenecientes a distintas categorías interactúen con blancos moleculares distintos garantiza cierta diversidad estructural, más aún si consideramos que aún los fármacos que pertenecen a una misma categoría terapéutica actúan en la mayoría de los casos por medio de mecanismos de acción diferentes. Por ejemplo: el enalapril y el losartán pertenecen a la misma categoría terapéutica (antihipertensivos), pero mientras el primero es un inhibidor de la enzima convertora de angiotensina el segundo es un antagonista de los receptores de angiotensina de tipo I. De manera similar, el amprenavir es un inhibidor de la proteasa del HIV mientras que el atevirdine y el abacavir son inhibidores de la transcriptasa reversa. Diferencias similares en los mecanismos de acción pueden encontrarse entre los fármacos de la categoría INACTIVA que pertenecen a otras categorías



terapéuticas, tales como los antibióticos y los antiulcerosos; b) se calcularon distintas distancias intermoleculares promedio de Tanimoto para todos los pares posibles de moléculas del conjunto de entrenamiento (siendo la distancia de Tanimoto igual a 1 menos el coeficiente de similaridad de Tanimoto) [8]. Las distancias promedio computadas fueron la distancia intramolecular promedio dentro de la categoría ACTIVA, la distancia intramolecular promedio dentro de la categoría INACTIVA y la distancia promedio entre las categorías ACTIVA e INACTIVA. A fin de realizar el cálculo de distancias se utilizaron tres conjuntos de subestructuras incluidas en el programa PowerMV del National Institute of Statistical Sciences de Estados Unidos [9]: pares de átomos (APs), pares de átomos de Carhart (CAPs) y huellas dactilares de farmacóforo (pharmacophore fingerprints - PFs). Los APs son subestructuras de la forma:



“Distancia” se refiere al número de átomos comprendidos entre los átomos X1 y X2 (los átomos que conforman el par) por el camino más corto, es decir, corresponde a la distancia topológica más 1. Por ejemplo, dos carbonos de hibridación sp^3 separados por una distancia topológica de 3 (esto es, a tres enlaces el uno del otro, por el camino más corto) generarían un AP del tipo:



Los CAP son subestructuras similares a la anterior sólo que, además de indicar la naturaleza de los átomos que componen el par, su estado de hibridación y la distancia topológica entre ambos indica



la cantidad de átomos distintos de hidrógeno a la que cada par del AP está unido.

Las FPs son más variadas que los pares de átomos. Por ejemplo, una subestructura considerada por PowerMV en la huella dactilar es **(dador de enlaces de puentes de H) - distancia topológica - (aceptor de enlaces de puentes de H)**. Otra es **(Anillo aromático - distancia topológica - Anillo aromático)**.

La tabla 6.2 muestra los valores de las distancias promedio de Tanimoto para cada uno de los tres conjuntos de subestructuras considerados. Teniendo en cuenta que dos estructuras idénticas tendrán una distancia de Tanimoto de 0 y dos estructuras sin subestructuras en común tendrán una distancia de Tanimoto de 1, indicado completa disimilitud, **las distancias intermoleculares promedio de la tabla indican que el conjunto de entrenamiento es, en efecto, estructuralmente diverso, independientemente de cuál de las subestructuras se utilice para el computo de la distancia**. Las estructuras de la categoría ACTIVA muestran una similaridad ligeramente mayor entre sí, aunque aún así la distancia promedio sigue siendo bastante alta; esto podría deberse a que interactúan con el mismo blanco molecular, lo que supone cierto patrón estructural común entre ellas. **Cabe destacar que la distancia promedio dentro de la categoría INACTIVA es casi idéntica a la distancia promedio entre la categoría ACTIVA y la categoría INACTIVA**. Es decir, los compuestos inactivos son tan disímiles entre sí como lo son de los compuestos activos.

2. **Cálculo del conjunto de D descriptores para el conjunto de entrenamiento.** Se calculó el valor de 877 descriptores de baja dimensionalidad para el conjunto de entrenamiento a través del programa Dragon (op. cit. en capítulo 3).
3. **Selección del conjunto de d descriptores a incluir en el modelo de entre los D descriptores totales. Generación del modelo.** Se aplicó



la técnica Stepwise Forward a distintos subconjuntos de los D descriptores (se trabajó sobre subconjuntos menores a los D descriptores totales, formados por distintas combinaciones de los grupos de descriptores de baja dimensionalidad que incluye Dragon, ya especificados en el capítulo 3. Se utilizó este criterio para minimizar la chance de correlación azarosa entre la variable dependiente y las variables independientes; por esta misma razón los programas de estadística no permiten utilizar números de descriptores demasiado grandes en un único análisis de MLR). Se trabajó con el módulo de análisis discriminante lineal del programa de estadística Statistica 7.0 [10], utilizando una tolerancia β de 0.5 para no incluir pares de descriptores

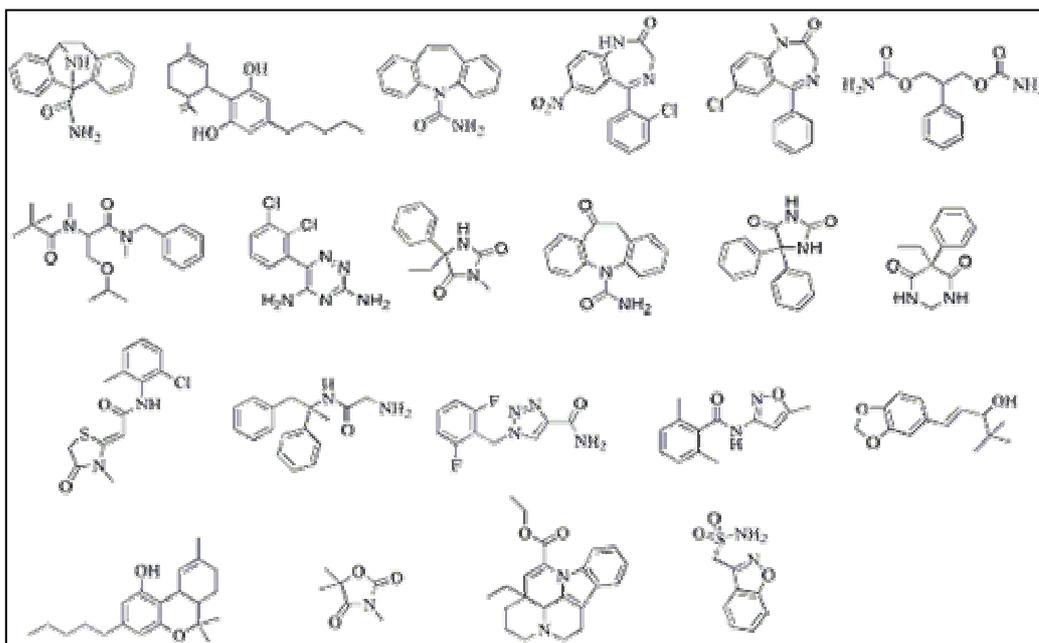


Fig. 6.1 – Estructuras de los 21 anticonvulsivos de uso clínico o en fase clínica de desarrollo que componen la categoría ACTIVA del conjunto de entrenamiento.

redundantes. Para elegir la mejor df de entre las funciones generadas se utilizaron los siguientes criterios: a) observación del % de buenas clasificaciones en el conjunto de entrenamiento; b) observación del valor del parámetro λ de Wilk (λ igual a 0 corresponde a una discriminación perfecta de las categorías consideradas, λ igual a 1 corresponde a capacidad de separación de



las clases nula) y; c) observación del principio de parsimonia: entre dos modelos con % de clasificación y λ similar se eligió el que incluyera menor número de descriptores.

El λ de Wilks es un estadístico que define si hay una diferencia significativa en la media de una combinación de variables para dos grupos previamente identificados; es una medida directa de la proporción de la varianza explicada por la

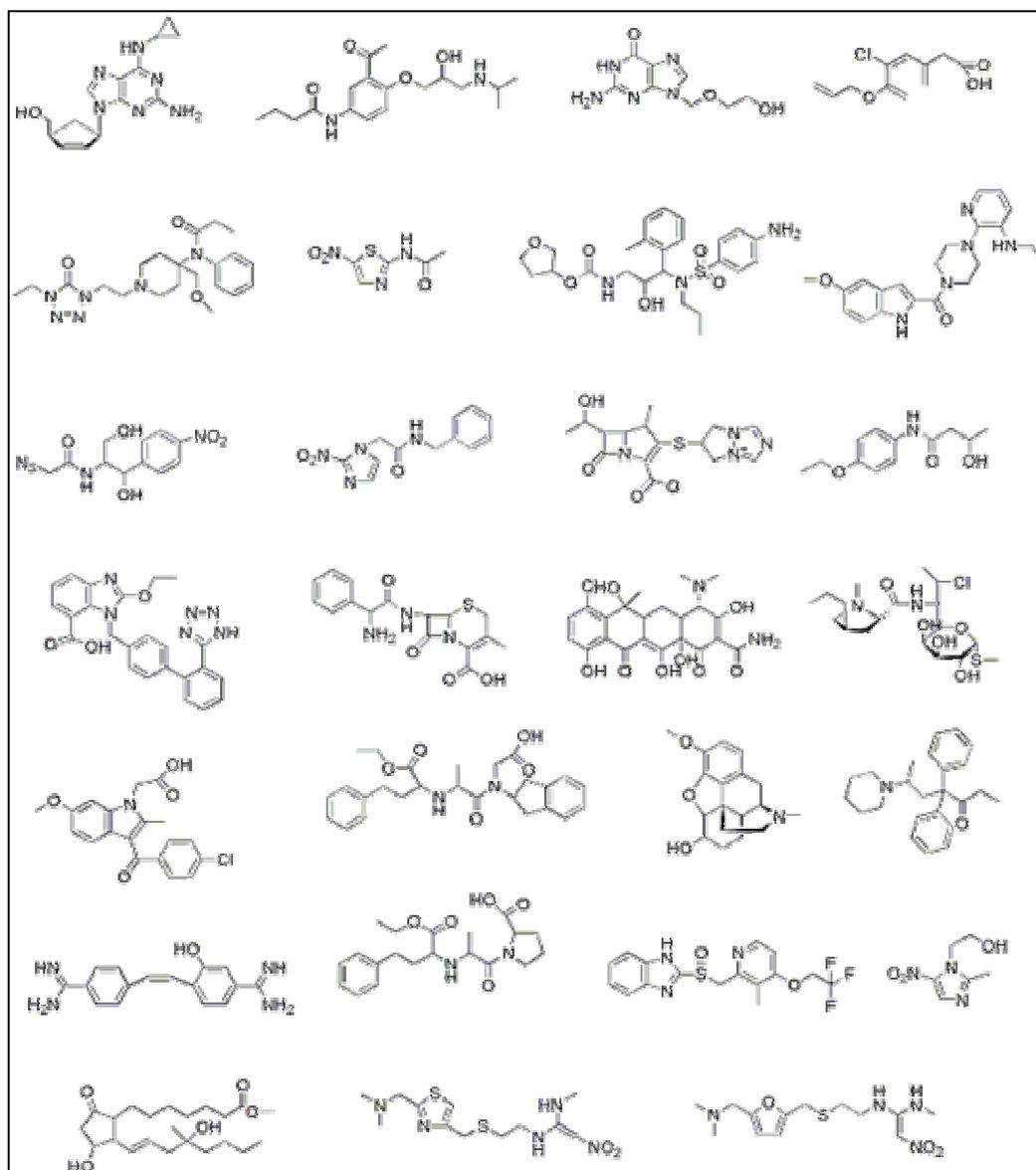


Fig. 6.2 – Estructuras de los 27 compuestos que integran la categoría INACTIVA del conjunto de entrenamiento.



variable clasificadora (es decir, evalúa la relación entre la varianza entre grupos y la varianza total).

Tabla 6.1. Compuestos que componen el conjunto de entrenamiento y categoría terapéutica de los mismos.

Compuesto	Categoría terapéutica	Compuesto	Categoría terapéutica
ADCI	Anticonvulsivo	Alclofenac	Analgésico, antipirético, antiinflamatorio
Canabidiol	Anticonvulsivo	Alfentanil	Analgésico (narcótico)
Carbamacepina	Anticonvulsivo	Aminitrozol	Antiprotozoario (tricomonas)
Clonazepam	Anticonvulsivo	Amprenavir	Antiviral
Diazepam	Anticonvulsivo	Ateviridine	Antiviral
Felbamato	Anticonvulsivo	Azidamfenicol	Antibacteriano
Lacosamida	Anticonvulsivo	Benzimidazol	Antiprotozoario (Tripanosoma)
Lamotrigina	Anticonvulsivo	Biapenem	Antibacterial
Mefenitoína	Anticonvulsivo	Bucetina	Analgésico
Oxcarbacepina	Anticonvulsivo	Candesartán	Antihipertensivo
Fenitoína	Anticonvulsivo	Cefalexina	Antibacteriano
Primidona	Anticonvulsivo	Clortetraciclina.	Antibacteriano
Ralitolina	Anticonvulsivo	Clindamicina	Antibacteriano
Remacemida	Anticonvulsivo	Clometacina	Analgésico
Rufinamida	Anticonvulsivo	Delapril	Antihipertensivo
Soretolide	Anticonvulsivo	Dihidrocodeína	Antitusivo
Estiripentol	Anticonvulsivo		Analgésico (narcótico)
THC	Anticonvulsivo	Dipipanona	Antihipertensivo
Trimetadiona	Anticonvulsivo	Enalapril	Antihipertensivo
Vinpocetina	Anticonvulsivo	Hidroxistilbamidina	Antiprotozoario (Leishmania)
Zonisamida	Anticonvulsivo	Lansoprazol	Antiulceroso
			Antiprotozoario (tricomonas), antiaméxico, antibacteriano
Abacavir	Antiviral	Metronidazol	Antiulceroso
Acebutolol	Antihipertensivo, antianginoso, antiarrítmico de clase II	Misoprostol	Antiulceroso
		Nizatidine	Antiulceroso
Aciclovir	Antiviral	Ranitidine	Antiulceroso



Tabla 6.2. Distancias intermoleculares promedio de los compuestos del conjunto de entrenamiento. Entre paréntesis, al lado de cada distancia promedio, se presenta la desviación estándar de las distancias intermoleculares. .

	Dentro de la categoría ACTIVA	Dentro de la categoría INACTIVA	Entre categorías
AP	0.73 (0.17)	0.63 (0.14)	0.67 (0.12)
CAP	0.79 (0.13)	0.80 (0.10)	0.82 (0.08)
FP	0.86 (0.23)	0.95 (0.09)	0.94 (0.09)

La variable dependiente utilizada asume, arbitrariamente, un valor de 1 para los compuestos de la categoría ACTIVA y un valor de -1 para los compuestos de la categoría INACTIVA. A continuación se presenta la df elegida:

$$df = 8.110 - 2.206 \times HVcpx - 4.277 \times BIC2 + 0.443 \times GATS7e + 1.089 \times GATS8p$$

$$\lambda \text{ de Wilks: } 0.32530 \quad F(4, 43)=22.297 \quad p < .0000 \quad N=48$$

HVcpx representa el índice de complejidad de los vértices del grafo [11], BIC2 simboliza el contenido de información de enlace (simetría de las vecindades de Segundo orden) [12], GATS7e denota la autocorrelación de Geary de orden 7, ponderada por las electronegatividades de Sanderson y GATS8p simboliza la autocorrelación de Geary de orden 8, pesada por las polarizabilidades atómicas [13].

Los valores de los coeficientes de regresión normalizados β son:

$$HVcpx \beta = -0.81; BIC2 \beta = -0.34; GATS7e \beta = 0.25; GARS8p \beta = 0.39.$$



de cuya comparación surge el peso relativo de cada descriptor en el modelo. Se observa que los cuatro descriptores contribuyen significativamente al valor de la *df*.

La Tabla 6.3 muestra la matriz de correlación de los cuatro descriptores incluidos; siguiendo el principio de no redundancia, todos los pares presentan una correlación baja aún aplicando criterios conservadores.

Tabla 6.3. Matriz de correlación de los descriptores incluidos en el modelo.

	HVcpx	BIC2	GATS7e	GATS8p
HVcpx	1.00	0.05	0.12	0.55
BIC2	0.05	1.00	-0.21	-0.02
GATS7e	0.12	-0.21	1.00	0.09
GATS8p	0.55	-0.02	0.09	1.00

La

relación casos/descriptores incluidos es más que adecuada (12, cuando la regla empírica establece que un número mayor 5 suele ser adecuado): el modelo cumple con todas las pautas especificadas en el capítulo 2. La Tabla 6.4 presenta los valores que asume la *df* para cada compuesto del conjunto de entrenamiento; puede verse que la misma clasifica correctamente el 93.75% de los compuestos del conjunto de entrenamiento. Las probabilidades de pertenencia a una categoría son estimadas en base a la distancia de Mahalanobis de cada compuesto al centroide de cada clase (el centroide es el punto del espacio N-dimensional, en este caso tetradimensional definido por los valores medios, para cada clase, de las variables independientes); la distancia de Mahalanobis es semejante a la distancia Euclidiana pero se utiliza en aquellos casos en los que las variables no son ortogonales (si las variables incluidas en el modelo no muestran correlación las distancias de Mahalanobis y Euclidiana coinciden).



Tabla 6.4. Valores de la *df* y probabilidad de pertenecer a la categoría activa para los compuestos del conjunto de entrenamiento. Los compuestos mal clasificados se indican con *.

Compuesto	df	Prob (Act)	Compuesto	df	Prob (Act)	Compuesto	df	Prob (Act)
ADCI	0.80	99.2	Estiripentol	0.17	73.9	Bucetina	-0.71	0.01
Canabidiol	0.33	88.0	THC	0.30	85.8	Candesartán	-0.97	0.00
Carbamacepina	1.21	99.9	Trimetadiona	0.79	99.1	Cefalexina	-0.82	0.01
Clonazepam	0.16	72.5	Vinpocetina	0.74	98.8	Clortetraciclina	-0.48	0.05
Diazepam	1.10	99.9	Zonisamida	0.32	87.2	Clindamicina	-1.05	0.00
Felbamato	0.96	99.7	Abacavir	-0.47	5.7	Clometacina	-0.70	0.01
Lacosamida	0.99	99.7	Acebutolol	-1.22	0.1	Delapril	-1.04	0.00
Lamotrigina	0.96	99.7	Aciclovir*	0.21	77.6	Dihidrocodeína*	0.50	0.95
Mefenitoína	0.51	95.5	Alclofenac	-0.83	0.7	Dipipanona*	0.65	0.98
Oxcarbacepina	0.75	98.9	Alfentanil	-1.01	0.2	Enalapril	-0.83	0.01
Fenitoína	1.39	100.0	Aminitrozol	-0.75	1.1	Hidroxistilbamidina	-0.16	0.28
Primidona	0.41	92.1	Amprenavir	-1.27	0.1	Lansoprazol	-0.99	0.00
Ralitolina	0.07	60.1	Atevirdine	-1.54	0.0	Metronidazol	-0.17	0.27
Remacemida	0.65	98.0	Azidamfenicol	-1.44	0.0	Misoprostol	-0.94	0.00
Rufinamida	0.01	52.0	Benznidazole	-0.62	2.3	Nizatidina	-1.00	0.00
Soretolide	0.68	98.3	Biapenem	-0.92	0.4	Ranitidina	-0.77	0.01

4. **Validación del modelo.** Se llevó a cabo la validación interna y externa de la *df*. La validación interna se realizó mediante la metodología LGO y ensayo de aleatorización. La metodología LGO se llevó a cabo extrayendo 6 compuestos del conjunto de entrenamiento por ronda LGO y agregando los compuestos extraídos al conjunto de prueba, regenerando el modelo con los 48-6 compuestos remanentes del conjunto de entrenamiento y observando el λ de Wilks y los porcentajes de clasificaciones acertadas en los conjuntos de entrenamiento y de prueba. Este procedimiento se repitió 8 veces. Los resultados se presentan en la tabla 6.5. Obsérvese que los modelos generados con 42 estructuras tienen un desempeño similar al modelo original (48 estructuras). El ensayo de aleatorización se llevó a cabo asignando aleatoriamente los valores de la variable dependiente a los compuestos del conjunto de entrenamiento y



regenerando la *df*. Este procedimiento se repitió 48 veces, tantas como número de compuestos integran el conjunto de entrenamiento. Los resultados pueden observarse en la fig. 6.3. Nótese que el modelo original es categóricamente superior a los modelos generados por aleatorización, indicando una probabilidad reducida de correlación al azar entre las variables independientes y la variable dependiente.

La **validación externa** se realizó sobre un conjunto de 48 compuestos: veinte compuestos activos en el ensayo MES en etapa de desarrollo pre-clínica extraídos de literatura [14-21] y 28 compuestos con otras bioactividades (antibióticos, antihipertensivos, antivirales, broncodilatadores, antiinflamatorios, antifúngicos, analgésicos) extraídos del Merck Index [6]. Los resultados se presentan en la tabla 6.6. El porcentaje de buenas clasificaciones es menor que el del conjunto de entrenamiento (75% contra 93.75%). Si bien es usual que el porcentaje de aciertos sea mayor en el conjunto a partir del cual se genera el modelo que en un conjunto independiente, vale la pena mencionar que el porcentaje de aciertos en el conjunto externo es mayor en el caso de la categoría INACTIVA del conjunto de prueba (aproximadamente del 79%). Es decir, pueden esperarse más falsos negativos que falsos positivos. El menor porcentaje de aciertos en la categoría ACTIVA del conjunto de prueba puede deberse a que se trata de compuestos disímiles a los de la categoría ACTIVA del conjunto de entrenamiento, debido a la decisión que hemos tomado de incluir únicamente compuestos de uso clínico o en fase clínica de evaluación en el conjunto de entrenamiento (en pos de seleccionar compuestos seguros y biodisponibles a nivel central tras administración oral); nótese que los compuestos anticonvulsivantes del conjunto de prueba no siempre han demostrado biodisponibilidad ni seguridad en humanos y que el carácter holístico de los



descriptores topológicos no sólo captura características que hacen a la interacción con el receptor sino también aspectos estructurales relacionados con la solubilidad, absorción a nivel intestinal, absorción a nivel de la BBB, velocidad de eliminación, toxicidad, etc. Es decir, la *df* selecciona compuestos no sólo por su actividad anticonvulsivante, y puede muy bien no seleccionar algunos de los compuestos en fase pre-clínica de evaluación por motivos distintos de su capacidad de interacción con el blanco molecular.

Dado al tamaño de la base de datos que se someterá a SV (cercana al medio millón de estructuras) los falsos negativos no nos preocupan mayormente: es de esperar que el número de compuestos seleccionados mediante SV exceda en mucho la capacidad de evaluación biológica. Si nos preocupa especialmente el número de falsos positivos,

Tabla 6.5. Resultados de la validación cruzada LGO. Las *dfs* obtenidas luego de la remoción al azar de seis compuestos en cada ronda LGO se desempeñan de manera casi idéntica a la *df* original

Compuestos removidos	N	λ de Wilks	% aciertos en conjunto de entrenamiento	% aciertos en conjunto de prueba
Ninguno (<i>df</i> original)	48	0.33	94	75
Clonazepam, lacosamida, oxcarbacepina, biapenem, candesartán, delapril	42	0.36	90	72
Rufinamida, THC, vinpocetina, benzimidazol, bucetina, clometacina	42	0.33	93	72
ADCI, mefenitoína, primidona, alfentanil, atevirdine, enalapril	42	0.35	90	74
Canabidiol, diazepam, fenitoína, aminitrozol, amprenavir, nizatine	42	0.35	93	74
Carbamazepina, felbamato, lamotrigina, acebutolol, cefalexina, dipipanona	42	0.30	93	72
Ralitolina, soretolide, estiripentol, azidamfenicol, lansoprazol, ranitidina	42	0.32	90	72
Remacemida, trimetadiona, zonisamida, alclofenac, clortetraciclina, dihidrocodeína	42	0.29	95	74
Clonazepam, lamotrigina, amprenavir, clindamicina, clometacina, misoprostol	42	0.35	90	80

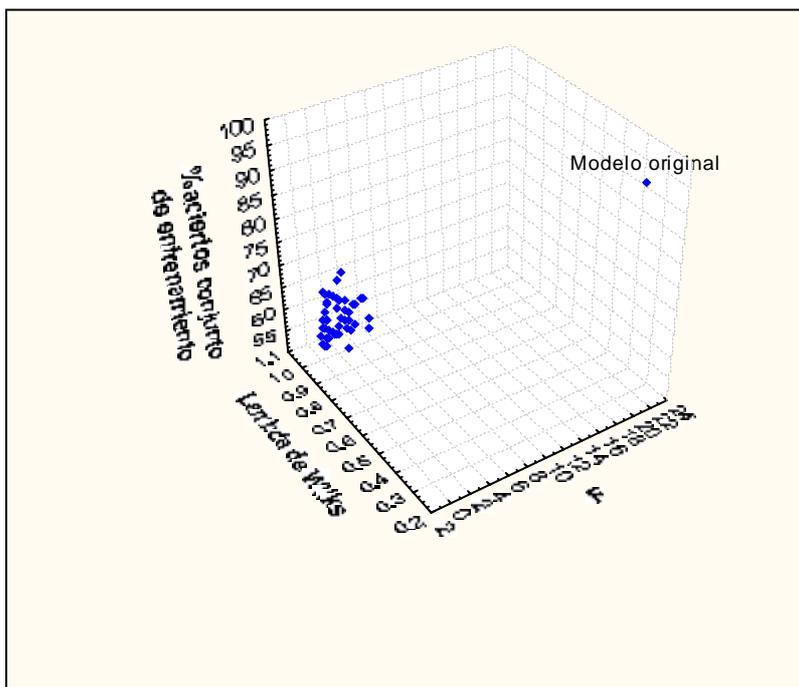


Fig. 6.3. Resultados del ensayo de aleatorización de Fisher.

por cuanto se desea evitar resultados negativos en los ensayos biológicos. Teniendo en cuenta estas consideraciones, nuestro criterio ha sido siempre minimizar el número de falsos positivos a expensas de aumentar el número de falsos negativos. También hemos aplicado este criterio, como se verá en breve, al utilizar un Diagrama de Distribución Farmacológica para elegir qué región de valores de la *df* se utilizará en el SV de la base de datos.

Tabla 6.6. Resultados de la validación externa. Los compuestos de la categoría INACTIVA se presentan escritos en *itálica*. Los asteriscos indican clasificaciones erróneas. La nomenclatura química se usa cuando no existe un nombre trivial.

Compuesto	Valor de <i>df</i>	Compuesto	Valor de <i>df</i>	Compuesto	Valor de <i>df</i>
534U87	0.26	<i>N-fenil-N_-(3,5-dimetilpirazol-4-il)urea</i>	0.55	<i>Carmoxirol</i>	-1.52
N-benzil-2-etilamino-3-metoxipropionamida *	-0.18	3,3,3-Trifluor-2-hidroxi-2-fenilpropionamida	0.48	<i>Ceftazol</i>	-1.58
N-bencil-2-(1-oxoetilamino)-2-fenilacetamida	0.05	3,3,3-Trifluor-2-hidroxi-N-metil-2-fenilpropionamida	0.96	<i>Clorobutanol</i> *	1.62

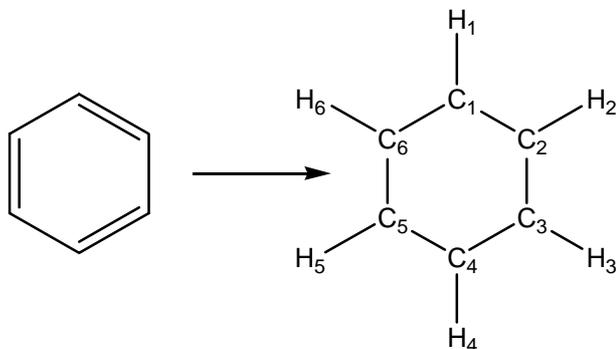


N-(2-fluorobencil)- 2- azaspiro[4.4]nonano - 1,3-diona	0.73	THIQ-10c *	-0.07	<i>Cimetidina</i>	-1.35
N-(3-fluorofenil)-2- azaspiro[4.5]decano- 1,3-diona	0.30	<i>Acetoxolona</i>	-0.30	<i>Cloxiquina</i>	-0.37
N-(3-trifluorometil fenil)-2-azaspiro [4.4]nonano- 1,3- diona	0.27	<i>Alacepril</i>	-1.37	<i>Cicloserina *</i>	0.45
CFM11	0.29	<i>Albuterol</i>	-0.02	<i>Desomorfina *</i>	0.20
CFM2 *	-0.13	<i>Amoxicilina</i>	-1.02	<i>Didanosina</i>	-0.22
CFM2S	0.20	<i>Aranidipina</i>	-0.03	<i>Dioxafetil Butirato *</i>	0.43
1- (4-metilpiperazin- 1-il) – 3 – (3- clorofenil) – pirrolidin-2,5- diona] *	-0.13	<i>Atenolol</i>	-1.09	<i>Doxofilina</i>	-0.95
1- (4-metilpiperazin- 1-il) – 3 – (4- clorofenil) – pirrolidin-2,5- diona	0.05	<i>Atorvastatina</i>	-1.01	<i>Ebrotidina</i>	-1.79
GYKI 53655	0.00	<i>Balofloxacina</i>	-0.48	<i>Efavirenz *</i>	0.54
NBQX *	-0.43	<i>Bambuterol</i>	-0.02	<i>Emtricitabina</i>	-0.12
4-amino-1,3,5- trimetilpirazol	0.68	<i>Benorilato</i>	-0.80	<i>Efedrina *</i>	0.21
<i>N-ciclohexil-N_- (3,5-dimetilpirazol- 4- il)tiourea</i>	0.20	<i>Bromfenac</i>	-0.44	<i>Ertapenem</i>	-2.09
<i>N-(4-metoxifenil)- N_-(3,5- dimetilpirazol- 4-il)tiourea*</i>	-0.54	<i>Capravirina</i>	-0.56		

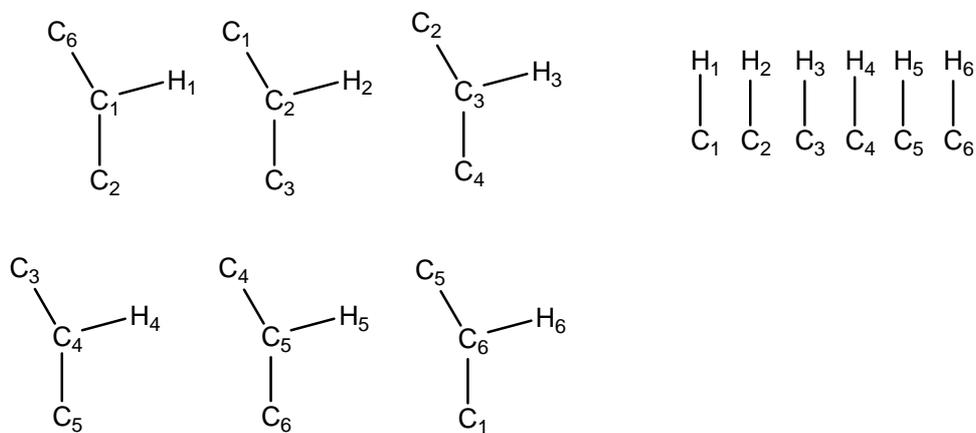
5. **Interpretación del modelo.** Pese a que los descriptores topológicos son característicamente abstractos, intentaremos una posible interpretación del modelo en los párrafos siguientes. HVcpx y BIC2 pertenecen a la clase de descriptores topológicos derivados de la teoría de la información, es decir, de la cuantificación mediante la Entropía de Shannon de la complejidad del conjunto de vecindades/subgrafos de un orden determinado que pueden derivarse del grafo molecular.



Considérese, por ejemplo, el grafo del benceno con hidrógenos explícitos.



El conjunto de vecindades de orden 1 que pueden obtenerse (cada vecindad integrada por un vértice del grafo y todos los otros vértices separados de aquel por una distancia topológica igual a 1) es:

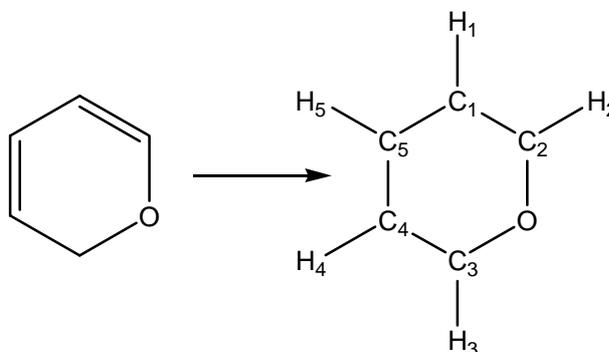


Y por lo tanto el Contenido de Información o Entropía de Shannon de primer orden, según se definió en el capítulo 3 puede calcularse como:

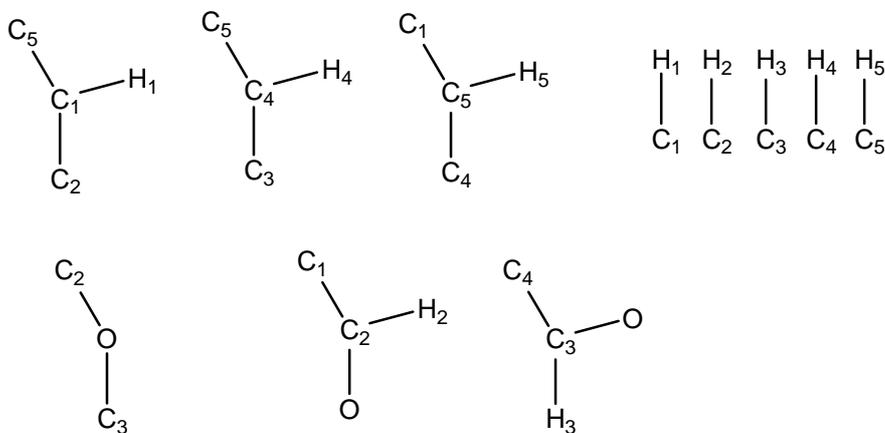
$$IC_1 = -(6/12 \log_2 6/12 + 6/12 \log_2 6/12) = 1.00$$



Considérese ahora, para una interpretación más acabada de la información reflejada por los Índices de Información, el grafo del 2H-pirano:



el conjunto de vecindades será en este caso:



y el contenido de información de primer orden:

$$IC_1 = -(5/11 \log_2 5/11 + 2/11 \log_2 2/11 + 1/11 \log_2 1/11 + 3/11 \log_2 3/11) = 1.79$$



Como puede verse, los índices de información caracterizan la complejidad del grafo molecular en base a la diversidad de los subgrafos de orden n que pueden derivarse del grafo. El hecho de que en la df elegida un incremento del valor de los dos índices de información incluidos sea desfavorable a la actividad anticonvulsiva podría estar reflejando el hecho de que los anticonvulsivos de la fig. 6.1 parecen ser estructuralmente menos complejos que los compuestos que conforman la categoría INACTIVA. Por ejemplo, los anticonvulsivos incluyen en promedio menos heteroátomos en su estructura, conduciendo a menor diversidad estructural en los conjuntos de vecindades que pueden derivarse. Podría argumentarse que existen, en las categorías terapéuticas incluidas en el grupo INACTIVO, compuestos menos complejos que los seleccionados. Tal es el caso por ejemplo de la aspirina, un analgésico-antiinflamatorio estructuralmente más simple que en principio podría haber sido incluido en el conjunto de entrenamiento. No obstante, se han encontrado reportes de que la aspirina posee actividad anti-MES [22,23].

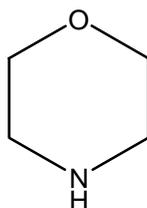
Las autocorrelaciones de Geary describen la distribución de una propiedad atómica determinada en la molécula. Se definen como:

$$GATSL(w) = \frac{1/2K \sum_i \sum_j (w_i - w_j) \delta(L, d_{ij})}{1/(n-1) \sum_i (w_i - \bar{w})^2}$$

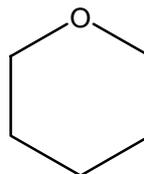
donde w_i representa el valor de la propiedad atómica considerada para el átomo i , L es el orden de la autocorrelación y $\delta(L, d_{ij})$ es el delta de Kronecker entre L y d_{ij} (lo que equivale a decir que sólo los pares de átomos separados por una distancia topológica igual a L serán contemplados en el cálculo de la autocorrelación). K es el número total de pares de átomos separados, en la estructura, por una distancia topológica L . La df incluye dos



autocorrelaciones de Geary considerando dos propiedades atómicas: electronegatividad y polarizabilidad, de orden 7 y 8 respectivamente. Según la df , la actividad anticonvulsiva anti-MES aumenta si GATS7e y GATS8p aumentan. Nótese que las autocorrelaciones de Geary asumen valores mayores cuando la molécula incluye, a una distancia topológica de L , átomos con diferencias en la propiedad atómica bajo consideración. La df podría entonces estar reflejando que: a) los anticonvulsivos, comparados con las estructuras INACTIVAS del conjunto de entrenamiento, tienden a poseer estructuras más compactas, con menor excentricidad, con menos cantidad de pares de átomos separados por distancias topológicas grandes (presentando por lo tanto menores valores de K para las autocorrelaciones de Geary de órdenes superiores) y; b) los anticonvulsivos tienden a presentar, en su estructura, una o a lo sumo dos regiones polares, con el resto de la molécula siendo más bien hidrofóbico, presentando por lo tanto un mayor número términos $(w_i - w_j)$ grandes comparados con otras estructuras con una distribución de carga más uniforme. En el caso de las estructuras inactivas algunos de esos términos pueden ser particularmente bajos por el hecho de que los heteroátomos parecen más uniformemente distribuidos en la molécula, habiendo una probabilidad mayor de encontrar átomos con electronegatividades o polarizabilidades similares a una distancia topológica de 7 u 8. Nuevamente presentamos un ejemplo con propósitos ilustrativos:



A



B



La autocorrelación de Geary de orden 3 pesada por electronegatividades atómicas GATS3v asumirá valores mayores en B que en A, porque la diferencia de electronegatividad entre el 0 y el C separados por una distancia topológica de 3 en B es mayor que la diferencia entre el N y el 0 separados por una distancia de 3 en A. Efectivamente, el valor de GATS3v es 0.878 para A y 0.972 para B.

6. **Validación final del modelo: evaluación biológica de estructuras predichas como anticonvulsivas.** Previamente a proceder a refinar el filtro generado (df) mediante la aplicación de un Diagrama de Distribución Farmacológica (DDF) y mediante adición de filtros secundarios, y antes de proceder al SV de las casi 460,000 estructuras químicas de la base de datos disponible, se realizó un SV sobre los compuestos del Merck Index 13th [6], con el fin de corroborar definitivamente que el modelo fuera capaz de identificar nuevos prototipos activos con actividad anticonvulsivante. Se realizó el SV de 10,250 compuestos (excluyendo compuestos inorgánicos). La tabla 6.7. presenta el valor de la df para 40 estructuras (20 clasificadas como activas y 20 clasificadas como inactivas) de entre las sometidas a SV. De las estructuras clasificadas como activas se eligieron 4 para su ensayo biológico en el MES test: ácido abiético, metilparabeno, abrine y acesulfame, obteniéndose en todos los casos resultados positivos (ver capítulo siguiente).

De acuerdo con los objetivos descritos en el capítulo 1, los criterios de selección de estas cuatro estructuras fueron: a) novedad estructural, b) información previa disponible respecto a su seguridad para administración en humanos y c) abundancia en la naturaleza o posibilidad de síntesis o hemisíntesis sencilla, para obtener fármacos de costo reducido y eventualmente aumentar la accesibilidad del paciente epiléptico al posible medicamento. En la



fig. 6.4 pueden observarse las estructuras seleccionadas. Nótese su diversidad estructural, que ratifica la capacidad del SV para identificar nuevos líderes (la distancia intermolecular promedio de Tanimoto entre las cuatro estructuras es de 0.705, 0.88 y 1.00 según se calcule en base a AP, CAP o PF).

Tabla 6.7. Resultados del SV para 40 de los compuestos del Merck Index 13th. Se presentan las categorías terapéuticas o usos (según Merck Index) y el valor de la *df* para cada compuesto. Los 20 de la mitad izquierda han sido clasificados como activos; con valores de la *df* por encima de 0. Los 20 restantes han sido clasificados como no anticonvulsivos.

Compuesto	Uso	<i>df</i>	Compuesto	Uso	<i>df</i>
Abrine	-	0.72	Abecarnil	Anxiolítico	-1.14
Ácido abiético	Manufatura de gomas	0.37	Abikoviromicina	Antiviral	-0.31
Acesulfame	Edulcorante	0.15	Acediasulfona	Antibacteriano	-0.26
Bufan	Insecticida	0.68	Acetamidoeugenol	Anestésico	-0.43
Difenidol	Antiemético	0.39	Buparvacuona	Antiprotozoario	-0.05
Demegestona	Progestogeno	0.43	Cilostazol	Antitrombótico	-1.40
Dropropizina	Antitusivo	0.03	Droxidopa	Antiparkinson	-0.47
Elenolida	-	0.39	Dulcin	Edulcorante	-0.46
Eltoprazina	Calmante	0.26	Ácido eicosapentanoico	Antihiperlipoproteínico	-1.40
Guanadrel	Antihipertensivo	0.19	Histamina	Antineoplásico	-0.23
Hexestrol	Antineoplásico	0.15	Hidralazina	Antihipertensivo	-0.24
Idazoxan	Antiparkinson	0.38	Hidroxianfetamina	Adrenérgico, midriático	-0.07
Idrocilamida	Relajante muscular	0.31	Levobunolol	Antiglaucoma	-0.30
Isosorbide	Diurético	0.68	Mefenesina	Relajante muscular	-0.11
Mequitazina	Antihistamínico	0.60	Nialamida	Antidepresivo	-1.39
Metapramina	Antidepresivo	0.64	Perazina	Antipsicótico	-0.08
Metilparabeno	Conservante	0.39	Succinilsulfatiazol	Antibacteriano	-1.36
Estanolona	Andrógeno	0.59	Zalcitabina	Antiviral	-0.11
Tetraetilfalamicida	Analéptico	0.74	Ácido zoledrónico	Inhibidor de la resorción ósea	-0.03
Zotepina	Antipsicótico	0.13	Zomepirac	Analgésico	-0.65

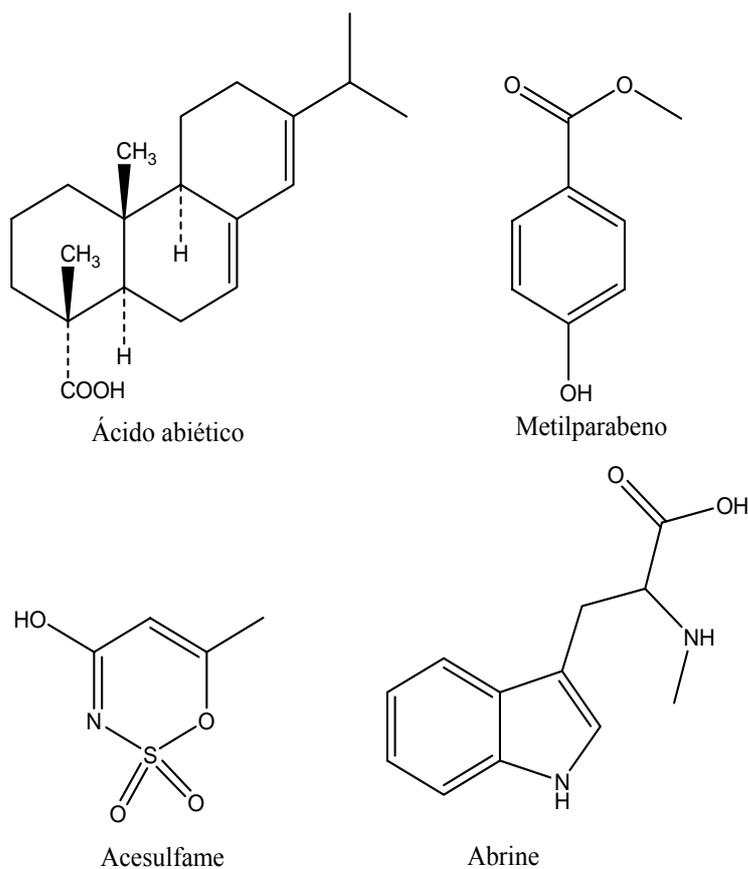


Fig. 6.4 Estructuras seleccionadas mediante SV del Merck Index 13th

7. **Aplicación del modelo en screening virtual de 459,718 estructuras químicas.** Antes del SV se construyó un Diagrama de Distribución Farmacológica para los compuestos del conjunto de entrenamiento [24]; en el mismo se grafica la expectativa de encontrar un compuesto inactivo o activo para diferentes intervalos de valores de la *df*. Dichas expectativas se calculan según:

$$\text{Expectativa activo} = \frac{\% \text{ de activos}}{\% \text{ de inactivos} + 100}$$

$$\text{Expectativa inactivo} = \frac{\% \text{ de inactivos}}{\% \text{ de activos} + 100}$$



donde el % de activos se refiere a:

$$\% \text{ de activos} = \frac{\text{número de activos en el intervalo considerado de la } df}{\text{número de activos totales utilizados para generar la } df} \times 100$$

y el % de inactivos se define de manera análoga:

$$\% \text{ de inactivos} = \frac{\text{número de inactivos en el intervalo considerado de la } df}{\text{número de inactivos totales utilizados para generar la } df} \times 100$$

El 100 en el denominador previene de dividir por 0 cuando no hay activos o inactivos, según el caso, en un intervalo de valores de la df determinado.

La figura 6.5 muestra el DDF obtenido. Puede verse que parte del área correspondiente a los compuestos inactivos invade el área correspondiente a los activos; esto refleja que la capacidad discriminante de la función no es perfecta (lo cual se evidencia a su vez en el valor del λ); aún así, la probabilidad de encontrar un compuesto inactivo en la zona de los activos es baja. En este caso, manteniendo la filosofía de disminuir el número de falsos positivos (compuestos que se enviarán a ensayo biológico sin resultados positivos) a costa de aumentar el número de falsos negativos (compuestos descartados a pesar de ser activos) se decidió seleccionar en el SV de las casi 460,000 estructuras sólo aquellos compuestos con valores de df entre 1.0 y 2.0, es decir, correspondientes a la zona de mínima superposición entre las dos categorías consideradas para el conjunto de entrenamiento.

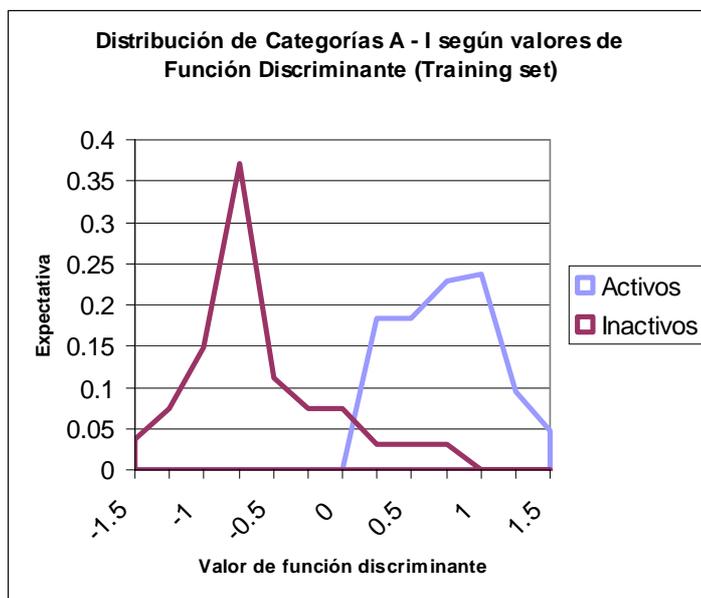


Fig. 6.5. DDF para el conjunto de entrenamiento. En base al diagrama, con la idea de reducir la incidencia de falsos positivos, se utilizará sólo la zona entre 1.0 y 2.0 en el SV.

Se procedió al cálculo de los 877 descriptores moleculares para 459,728 estructuras extraídas de la base ZINC database [25]. Si bien el proceso de SV se hubiera acelerado de calcular únicamente los cuatro descriptores incluidos en la df , se prefirió calcular todos los descriptores de baja dimensionalidad del Dragon con el fin de facilitar futuros estudios de SV orientados a otros blancos moleculares; nótese que otros miembros del grupo de trabajo han conducido estudios de este tipo, con posterioridad, para la búsqueda de nuevos antichagásicos y nuevos antiepileptogénicos [26, 27].

Una vez calculados los descriptores, se identificaron las estructuras de la base de datos con valores de la df en el intervalo [1.0-2.0]. Se identificaron 2,719 estructuras potencialmente anticonvulsivas en el ensayo MES mediante aplicación de este criterio y los filtros generales de absorción que se describen en el capítulo 7.



Referencias

1. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. A novel approach for the virtual screening and rational design of anticancer compounds. *J. Med. Chem.* **2000**, 43, págs. 1975-1985
2. Gálvez, J.; García-García, A.; de Julián Ortiz, J.V.; García-Domenech, R.; Muñoz, C.; Gunal, R.; Borrás, R. New agents active against *Mycobacterium avis* complex selected by molecular topology: a virtual screening method. *J. Antimicrob. Chemother.* **2004**, 53, págs. 65-73
3. Julián-Ortiz, J.V.; Gálvez, J.; Muñoz, C.; García-Domenech, R.; Gimeno-Cardona, C. Virtual combinatorial synthesis and computational screening of anti-herpes compounds. *J. Med. Chem.* **1999**, 42, págs. 3308-3314
4. Meneses-Marcel, A.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Montero-Torres, A.; Montero-Pereira, D.; Escario, J.A.; Nogal-Ruiz, J.J.; Ochoa, C.; Arán, V.J.; Martínez-Fernández, A.R.; García Sánchez, R.N. A linear discriminant analysis based virtual screening of trichomonacidal lead-compounds: outcomes of in silico studies supported by experimental results. *Bioorg. Med. Chem. Lett.* **2005**, 15, págs. 3838-3843
5. Montero-Torres, A.; García-Sánchez, R.N.; Marrero-Ponce, Y.; Machado-Tugores, Y.; Nogal-Ruiz, J.J.; Martínez-Fernández, A.R.; Arán, V.J.; Ochoa, C.; Meneses-Marcel, A.; Torrens, F. Non-stochastic quadratic fingerprints and LDA-based QSAR models in hit and lead generation through virtual screening: theoretical and experimental assessment of a promising method for the discovery of new antimalarial compounds. *Eur. J. Med. Chem.* **2006**, 41, págs. 483-493
6. The Merck Index, Encyclopedia of Chemicals, Drugs and Biologicals, 13ra edición. Merck and Co, New Jersey, Estados Unidos, **2001**.
7. National Institute of Health, Pubchem, <http://pubchem.ncbi.nlm.nih.gov/>
8. Karakoc, E.; Cherkasov, A.; Cenk Sahinalp, S. Distance based algorithms for small biomolecule classification and structural similarity search. *Bioinformatics.* **2006**, 22, págs. 243-351
9. National Institute of Statistical Sciences. PowerMV version 0.61. **2005**. <http://www.niss.org/PowerMV>
10. Statsoft, Inc. STATISTICA version 7.0, **2004**. <http://www.statsoft.com>
11. Raychaudhury, C.; Ray, S.K.; Ghosh, J.J.; Roy, A.B.; Basak, S.C. Discrimination of isomeric structures using information theoretic topological indices. *J. Comput. Chem.* **1984**, 5, págs. 581-588
12. Magnuson, D.R.; Harris, V.K.; Basak, S.C. en King, R.B. (editor), *Studies in Physical and Theoretical Chemistry*. Elsevier, Amsterdam, Holanda, **1983**, págs. 178-191



13. Geary, R.C. The Contiguity Ratio and Statistical Mapping. *Incorp Statist.* **1954**, 5, págs. 115-145
14. Obniska, J.; Kaminski, K.; Zagorska, A.; Dzierzawska-Majewska, A.; Karolak-Wojciechowska, J. Synthesis and anticonvulsant activity of new fluorinated N-phenyl- and N-benzyl-2-azaspiro[4.4]nonane- and [4.5]decane-1,3-dione derivatives: Part III. *J. Fluorine Chem.* **2006**, 127, págs. 417-425
15. Bialer, M.; Johannessen, S.I.; Kupferberg, H.J.; Levy, R.H.; Loiseau, P.; Perucca, E. Progress report on new antiepileptic drugs: a summary of the fourth Eilat conference (EILAT IV). *Epilep. Res.* **1999**, 34, págs. 1-41
16. De Sarro, G.; Ferreri, G.; Gareri, P.; Russo, E.; De Sarro, A.; Gitto, R.; Chimirri, A. Comparative anticonvulsant activity of some 2,3-benzodiazepine derivatives in rodents). *Pharmacol., Biochem. Behav.* **2003**, 74, págs. 595-602
17. Ferreri, G.; Chimirri, A.; Russo, E.; Gitto, R.; Gareri, P.; De Sarro, A.; De Sarro, G. Comparative anticonvulsant activity of N-acetyl-1-aryl-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline derivatives in rodents *Pharmacol., Biochem. Behav.* **2004**, 77, págs. 85-94
18. Obniska, J.; Jurczyk, S.; Zejc, A.; Kamiński, K.; Tatarczyńska, E.; Stachowicz, K. Anticonvulsant properties of N-(4-methylpiperazin-1-yl)- and N-[3-(4-methyl-piperazin-1-yl)propyl] derivatives of 3-aryl- and 3-spirocycloalkyl-pyrrolidine-2,5-dione. *Pharmacol. Rep.* **2005**, 57, págs. 170-175
19. Schenck, H.A.; Lenkowski, P.W.; Choudhury-Mukherjee, I.; Ko, S.; Stables, J.P.; Patel, M.K.; Brown, M.L. Design, synthesis and evaluation of novel hydroxyamides as orally available anticonvulsants. *Bioorg. Med. Chem.* **2004**, 12, págs. 979-993
20. Beguin C.; LeTiran A.; Stables J. P.; Voyksnerc R. D.; Kohnd H. N-Substituted amino acid N'-benzylamides: synthesis, anticonvulsant, and metabolic activities. *Bioorg. Med. Chem.* **2004**. 12, págs. 3079-3096
21. Kaymakçioğlu, B.K.; Rollas, S.; Korcegez, E.; Arıcıoğlu, F. Synthesis and biological evaluation of new N-substituted-N'-(3,5-di/1,3,5-trimethylpyrazole-4-yl)thiourea/urea derivatives. *Eur. J. Pharm. Sci.* **2005**, 26, págs. 97-103
22. Wali, R.S.; Patil, P.A. Aspirin and anticonvulsant interaction. *Indian J. Physiol. Pharmacol.* **1995**, 39, págs. 77-79
23. Srivastava, A.K.; Gupta, Y.K. Aspirin modulates the anticonvulsant effect of diazepam and sodium valproate in pentylenetetrazole and maximal electroshock induced seizures in mice. *Indian J. Physiol. Pharmacol.* **2001**, 45, págs. 475-480
24. Gálvez, J.; García-Domenech, R.; Alapont de Gregorio, C.; de Julián-Ortiz, J.V.; Popa, L. Pharmacological Distribution Diagrams:



- a tool for de novo drug design. *J. Mol. Graph.* **1996**, 14, págs. 272-276.
25. Irwin, J.J.; Shoichet, B.K. ZINC - A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Comput. Sci.* **2005**, 45, págs. 177-182
26. Prieto, J.J.; Talevi, A.; Bruno-Blanch, L.E. Application of linear discriminant analysis in the virtual screening of antichagasic drugs through trupanothione reductase inhibition. *Mol. Divers.* **2006**, 10, págs. 361-375
27. Bellera, C.L.; Talevi, A.; Bruno-Blanch, L.E. Aplicación de análisis lineal discriminante en la búsqueda de drogas antiepileptogénicas. *Latin American J. Pharm.* **2007**, 26, págs. 244-252



Capítulo 6. Ensayos biológicos.

“De lo que oigas no creas nada; de lo que veas, cree la mitad”

Dicho popular

“La Fe está en el que duda”

Dalmiro Saénz y Alberto Cormillot, *Cristo de Pie*

Con excepción del fenobarbital y los bromuros, la mayoría de los fármacos anticonvulsivos utilizados en clínica han sido identificados mediante modelos animales [1]. Los modelos animales de epilepsia pueden dividirse en dos grandes grupos: **modelos agudos** y **modelos crónicos**. En el caso de los modelos agudos, las convulsiones se inducen artificialmente en animales sanos mediante agentes físicos (electricidad, temperatura, sonido, luz) o químicos (pentilentetrazol, kainato, pilocarpina, picrotoxina, estriquina, etc.). En los modelos crónicos se induce en cambio una susceptibilidad a desarrollar convulsiones, ya sea mediante administración repetida de estímulos convulsivos en dosis sub-convulsivas o mediante el uso de animales con susceptibilidad genética a desarrollar epilepsia (ya sea que se trate de una línea de animales que espontáneamente desarrolle epilepsia, como la línea de ratas GEPRS o de líneas mutantes o transgénicas obtenidas por medio de la biotecnología) [1,2].

El MES test es el modelo animal que ha sido más utilizado en el descubrimiento de nuevos agentes antiepilépticos [2], debido a que la inducción de la convulsión es simple y el valor predictivo para detectar fármacos antiepilépticos clínicamente efectivos es alto. El



ensayo MES ha sido correlacionado con las convulsiones tónico-clónicas generalizadas [3], y se ha propuesto que también puede estar vinculado a las convulsiones parciales [2]. Sin embargo, el MES test falla en identificar anticonvulsivos efectivos en clínica cuyo mecanismo de acción se relaciona con el aumento de la transmisión GABAérgica, tales como el vigabatrin y la tiagabina, anticonvulsivos ambos eficaces contra las convulsiones parciales en humanos. El MES test no sólo da lugar a falsos negativos sino también a falsos positivos: los bloqueantes de receptores de glutamato protegen de las convulsiones en el MES test, pero no han demostrado eficacia clínica en el control de la epilepsia. Esto podría deberse a la acción relajante muscular de estas drogas; el MES test es muy sensible a relajantes musculares, lo que evidencia la importancia de realizar, concomitante al ensayo MES, otros ensayos que verifiquen que la droga administrada no produce ataxia. Este comportamiento evidencia las ventajas y limitaciones del ensayo MES.

Otro ensayo ampliamente utilizado es el ensayo PTZ, que identifica drogas efectivas contra convulsiones mioclónicas, mediante inyección subcutánea del convulsivante pentilentetrazol. Una de las desventajas de este ensayo es que no permite detectar fármacos activos contra crisis epilépticas sin convulsión, como ser la lamotrigina. Puede decirse que a la fecha no se ha desarrollado aún un modelo animal de epilepsia único sensible a todos los fármacos anticonvulsivantes conocidos [2]. Por lo tanto, hasta el momento los modelos animales disponibles deben utilizarse en forma complementaria en la fase temprana de desarrollo de nuevos antiepilépticos.

En el presente trabajo hemos evaluado los cuatro candidatos identificados mediante SV del Merck Index 13th en el ensayo MES. En algunos casos también se llevó a cabo el ensato PTZ. Se han seguido



procedimientos estándar indicados por el programa del Instituto Nacional de Salud de Estados Unidos (NIH) [4], además de respetarse las pautas internacionales para el uso de animales en ensayos biomédicos (Council for International Organizations of Medical Sciences, op. cit. capítulo 4). La presencia de actividad en el metilparabeno motivó el ensayo del propilparabeno, agente conservante ampliamente utilizado, conjuntamente con el metilparabeno, en productos cosméticos, farmacéuticos y alimenticios (ver fig. 7.1).

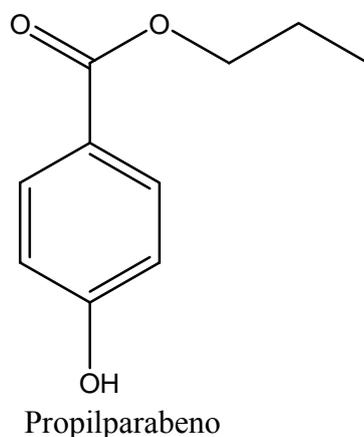


Fig. 7.1. Estructura química del propilparabeno.

Para todos los ensayos realizados se utilizaron ratones Swiss de entre 18 y 23 gramos. Los ratones se guardan en jaulas para colonias con ciclos de luz/oscuridad de 12 horas; se los provee con agua y comida *ad libitum*. Los compuestos se disuelven en agua, administrándose un volumen máximo de 10 ml de solución recientemente preparada por kg de ratón. Cuando la droga no era soluble en agua se utilizó hasta un 30% de PEG 400. El Programa de Desarrollo de Nuevos Fármacos Antiepilépticos del NIH prevé la administración, en su primera fase, de tres dosis de la droga a evaluar: 30, 100 y 300 mg/kg, evaluándose la actividad a 0.5 y 4 hs luego de la administración. Se considera promisorio a un compuesto activo que



protege de la convulsión inducida a 30 ó 100 mg/kg, debiendo continuarse con las siguientes fases de evaluación, en las que se evalúa otra especie animal y se determinan parámetros tales como el tiempo de efecto máximo, la ED₅₀ y la TD₅₀ y el índice de protección en ratones y ratas.

En el ensayo MES se administró al ratón un estímulo eléctrico de 60 Hz/50 mA durante 0.2 segundos mediante electrodos de oreja. Una gota de solución salina o gel para electroforesis fueron colocados en las orejas del animal antes del electrodo para asegurar contacto eléctrico adecuado. En estas condiciones se producen convulsiones en virtualmente todos los ratones. La convulsión típicamente consiste en un corto período de flexión tónica seguido de un período más prolongado de extensión tónica de los miembros posteriores y un episodio clónico final. Se considera que la abolición de la fase de extensión tónica indica que la droga administrada protege de la convulsión (los miembros posteriores no deben formar un ángulo de más de 90 grados con el tronco).

En el caso del ensayo PTZ, luego de administrado el fármaco a evaluar se inyecta, a los dos tiempos estipulados, una solución de PTZ al 1.7% en solución salina 0.9% dentro de un pliegue de piel en la línea media del cuello, en un volumen de 5 ml/kg de peso. Los animales se observan durante 30 minutos luego de la inyección para determinar presencia o ausencia de un episodio convulsivo de duración igual o mayor a 5 segundos. La ausencia de un episodio clónico indica protección.

Se realizó además, en todos los casos, el ensayo del Rotorod, para verificar que la droga inyectada no produzca ataxia dando lugar a un falso positivo. Para ello, los animales ensayados se colocan en un cilindro que rota a 6 rpm. La incapacidad para mantener el equilibrio



durante 1 minuto se toma como evidencia de que el fármaco administrado ha producido falta de coordinación motora.

La tabla 7.1 presenta los resultados de los tres ensayos para las cinco drogas ensayadas. Todas las drogas presentaron actividad a 30 mg/kg. Cabe destacar que ninguna droga presentó actividad en el ensayo PTZ; esto es consecuencia de que, debido a que el objetivo de esta tesis se relaciona con la identificación de nuevos anticonvulsivos activos en el ensayo MES, en el conjunto de entrenamiento de la *df* sólo se incluyeron drogas con actividad anti-MES, evitando la inclusión de fármacos con actividad anti-PTZ tales como vigabatrina, ácido valproico y tiagabina. Por otro lado, ninguna de las drogas produjo ataxia.

Tabla 7.1. Resultado de los ensayos biológicos de las drogas seleccionadas; se indica el número de animales en los que se observó protección sobre el número de animales totales evaluados.

Droga	Dosis (mg/kg)	Vía de administración	MES test		PTZ test		Rotorod test	
			0.5 hs	4 hs	0.5 hs	4 hs	0.5 hs	4 hs
Ácido abiético	30	i.p.	0/3	4/8	0/2	0/2	0/3	0/17
		p.o.	-	2/5	-	-	-	-
	60	p.o.	-	1/3	-	-	-	0/3
	100	i.p.	0/3	2/8	0/2	0/2	0/3	0/17
	200	p.o.*	-	1/5	-	-	-	0/3
		p.o.*	-	1/3	-	-	-	0/3
Metilparabeno	30	i.p.	1/3	1/3	-	-	0/3	0/3
	100	i.p.	2/3	0/3	-	-	0/3	0/3
	300	i.p.	1/3	1/3	-	-	0/3	0/3
Propilparabeno	30	i.p.	2/3	1/3	0/3	0/3	0/6	0/6
	100	i.p.	2/3	1/3	0/3	0/3	0/6	0/6
	300	i.p.	1/3	3/3	0/3	0/3	0/6	0/6
Acesulfame	30	i.p.	0/3	1/3	0/2	0/2	0/5	0/5
	100	i.p.	0/3	1/3	0/2	0/2	0/5	0/5
	300	i.p.	2/3	2/3	0/2	0/2	0/5	0/5
Abrine	30	p.o.*	1/3	0/3	0/2	0/2	0/5	0/5
	100	p.o.*	1/3	0/3	0/2	0/2	0/5	0/5
	300	p.o.*	0/3	0/3	0/2	0/2	0/5	0/5

* administrado como suspensión



Basados en los resultados anteriores se determinó, adicionalmente, el tiempo de efecto máximo del propilparabeno, inyectando 100 mg/kg (i.p.) a tiempos 0.25, 0.5, 1, 2, 3 y 4 horas, obteniendo protección, en ese orden, en 0/3, 1/3, 0/3, 2/3, 1/3 y 1/3 animales evaluados. Se tomó $t = 2$ hs como el tiempo de efecto máximo. Luego se procedió a la determinación de la ED_{50} del propilparabeno, para lo cual se inyectaron dosis de 30, 80, 130 y 180 mg/kg y se evaluó la actividad anticonvulsiva a las dos horas, obteniéndose protección en 1/5, 1/5, 3/5 y 4/5 animales, respectivamente. La ED_{50} calculada graficando el PROBIT contra el logaritmo de la dosis administrada es de 101.7 mg/kg.

Se observa que los fármacos ensayados poseen un perfil anticonvulsivo interesante, en especial en los casos del propilparabeno, el acesulfame y el ácido abiético.

Tanto el acesulfame como los parabenos han sido ampliamente utilizados, desde hace años, como edulcorante y conservantes, en ese orden, sin haber mostrado toxicidad aguda o crónica. El metil y propilparabeno han demostrado seguridad en estudios de administración crónica en dosis mucho mayores a las administradas en el presente trabajo [5,6]. No se observó carcinogenicidad ni toxicidad a largo plazo por administración de dosis de entre 0.9-1.2 y 5.5-5.9 g/kg/día de metilparabeno y propilparabeno, respectivamente, en ratas durante 96 semanas. La administración de 550 mg/kg/día de metilparabeno a ratones y ratas preñados no tuvo efecto sobre la nidación, la tasa de mortandad de fetos y de crías; tampoco indujo anormalidades en el tejido esquelético. Ninguno de los parabenos mostró evidencia de mutagénesis en el test de Ames. El hecho de disponer de datos de seguridad y farmacocinética bien documentados de algunos de los



compuestos ensayados podría facilitar el desarrollo de medicamentos seguros a partir de los mismos, lo que como ya se dijo constituye una de las ventajas del SV: identificar nuevas actividades en compuestos con perfiles toxicológico y farmacológico conocidos, permitiendo responder con mayor prontitud a necesidades urgentes de la población enferma. Teniendo en cuenta que algunos de los compuestos ensayados se usan habitualmente en formulaciones pediátricas de anticonvulsivos (aunque en dosis bastante menores a las aquí evaluadas) sería conveniente evaluar en un futuro si existe sinergismo entre la acción anticonvulsiva de estos compuestos y la de anticonvulsivos actualmente utilizados en clínica.

Referencias

1. Cole, A.C.; Koh, S.; Zheng, Y. Are seizures harmful: what can we learn from animal models. *Prog. Brain Res.* **2002**, 135, págs. 13-23
2. Löscher, W. Animal models of epilepsy for the development of antiepileptogenic and disease-modifying drugs. A comparison of the pharmacology of kindling and post-status epilepticus models of temporal lobe epilepsy. *Epilepsy Res.* **2002**, 50, págs. 105-123
3. White, H.S. Clinical significance of animal seizure models and mechanism of action studies of potential antiepileptic drugs. *Epilepsia.* **1997**, 38 Suppl. 1, págs. S9-S17.
4. Stables, J. P.; Kupferberg, H. J.; Gladding, R. Molecular and cellular targets for antiepileptic drugs. In Avanzini, G.; Regesta, G.; Tanganelli, P.; Avoli, M.; (editors), John Libbey, Londres, **1997**.
5. Soni, M.G.; Taylor, S.L.; Greenberg, N.A.; Burdock, G.A. Evaluation of the health aspects of methyl paraben: a review of the published literature. *Food Chem. Toxicol.* **2002**, 40, págs. 1335-1373
6. Soni, M.G.; Burdock, G.A.; Taylor, S.L.; Greenberg, N.A. Safety assessment of propyl paraben: a review of the published literature. *Food Chem Toxicol.* **2001**, 39, 513-532



Capítulo 7. Filtros ADME.

“Ser y no estar es no ser y estar sin ser no vale la pena”

Dalmiro Sáenz, *El Depredador*

- Filtros ADME/Tox generales.
- Filtros ADME/Tox particulares.

Como ya se mencionó, el costo creciente del desarrollo de nuevos fármacos se debe en parte a que recién cuando los compuestos se están evaluando en las fases clínica I, II y III, surgen los problemas vinculados a su posible uso como medicamentos (justamente, las fases más costosas en un programa de investigación y desarrollo de nuevos fármacos). Los problemas en cuestión incluyen características no óptimas de absorción, distribución, metabolismo y excreción, así como también toxicidad, aspectos a los que nos referimos usualmente como ADME/Tox o ADMET. Como consecuencia de esta situación, las industrias farmacéuticas internacionales han adoptado una estrategia que podría sintetizarse bajo el lema “si fallas temprano, fallas barato” [1-4]. No podemos insistir lo suficiente en el hecho de que existen fuertes cuestiones éticas que obligan al investigador a realizar cuantos esfuerzos estén a su alcance para disminuir la tasa de fracasos en las inevitables etapas pre-clínicas y clínicas de evaluación: en ambos casos se involucran seres vivos en los estudios, y en el segundo caso, seres humanos.



Es así que un gran número de ensayos *in vitro* han sido desarrollados para evaluar las propiedades de absorción (por ej. los modelos con células Caco-2 y MDCK) y la metabolización utilizando microsomas o hepatocitos [5]. Si bien estos modelos *in vitro* son sumamente valiosos, los mismos tienen algunas desventajas: requieren muestras para realizar los ensayos, y consumen recursos y tiempo (estos últimos no siempre son compatibles con los tiempos logrados en el screening de alta performance). Esto ha motivado un gran interés en el desarrollo de filtros computacionales para predecir las características ADMET de compuestos activos.

De acuerdo con la filosofía anteriormente expuesta, en el presente estudio se han aplicado, conjuntamente con la *df* ya descrita, **filtros ADME/Tox generales** (aquellos que no fueron desarrollados específicamente para este trabajo de tesis, siendo de aplicación extendida en el campo de desarrollo de nuevos fármacos a nivel de la industria farmacéutica internacional) y **dos filtros ADME/Tox particulares**: un modelo de permeación a través de la BBB y un segundo filtro referido a la neurotoxicidad de los compuestos evaluados.

Filtros ADME/Tox generales.

Se utilizaron dos filtros ADME/Tox generales: la llamada “**regla de cinco de Lipinski**” y el **valor óptimo del coeficiente del log P para difusión pasiva a través de la BBB**.

La regla de cinco de Lipinski es una regla empírica derivada del análisis de bases de datos de fármacos de uso clínico (sin considerar polipéptidos y sales de amonio cuaternario) [6]. La misma establece que una pobre permeabilidad o absorción de un fármaco al ser



administrado por vía oral es muy probable si dos o más de las siguientes condiciones no se cumplen:

- Peso molecular menor a 500
- Número de dadores de enlaces de hidrógeno menor o igual a 5
- Número de aceptores de enlaces de hidrógeno menor o igual a 10
- Coeficiente de reparto octanol-agua entre -5.0 y 5.0 (equivalente a $\log P$ de Moriguchi $-\text{mlogP}$ entre -4.15 y 4.15).

La regla de Lipinski expresa, mediante una serie de condiciones sencillas, el hecho conocido de que la biodisponibilidad por vía oral de un fármaco depende (además de cuestiones relacionadas con la tecnología farmacéutica del vehículo) de un balance entre su hidrofilia y su lipofilia. En primer lugar, el fármaco debe estar en solución para poder absorberse y la velocidad de absorción a través de la barrera intestinal será por otro lado aproximadamente proporcional a la diferencia de concentraciones del fármaco a ambos lados de la membrana intestinal. Sin embargo, un fármaco excesivamente polar presentará dificultades de permeabilidad a través de las membranas biológicas. Las reglas de Lipinski son uno de los criterios más usados en la actualidad para definir si un compuesto químico es “drug-like”, es decir, si posee una serie de características estructurales que lo convierten, potencialmente, en un fármaco a partir del cual se puede desarrollar un medicamento [7]. En los histogramas de la Figs. 8.1 a 8.4 puede observarse que casi la totalidad de las estructuras de la base de datos de 459,718 estructuras cumplen con la regla de cinco. Hemos analizado los valores de los cuatro descriptores involucrados en la regla de Lipinski para las estructuras de la base de datos, determinando que el 99.3% de las mismas cumple con al menos tres de las condiciones de la regla de Lipinski y por lo tanto presentan altas probabilidades de presentar buena biodisponibilidad oral (para ello se



ha restado al número total de estructuras el número de estructuras que no cumplen con dos, tres y cuatro de las condiciones de la regla).

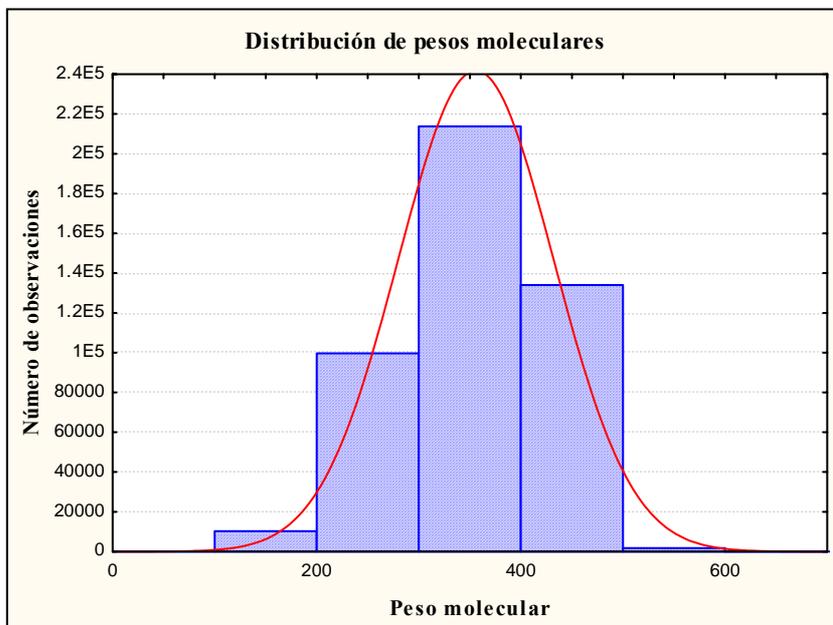


Fig. 8.1. Distribución de pesos moleculares de las 459,718 estructuras de la base de datos.

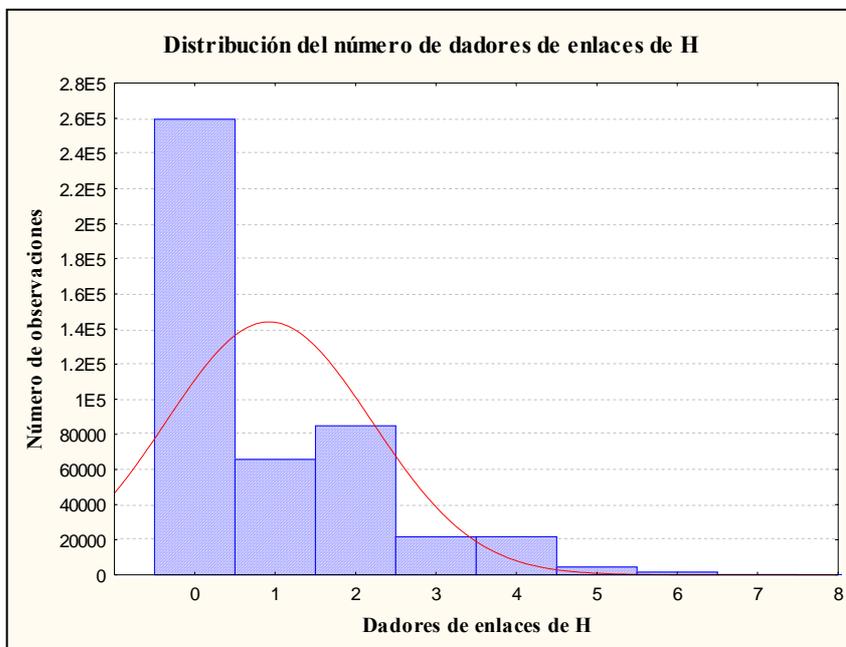


Fig. 8.2. Distribución del número de dadores de enlaces de H para las 459,718 estructuras de la base de datos.

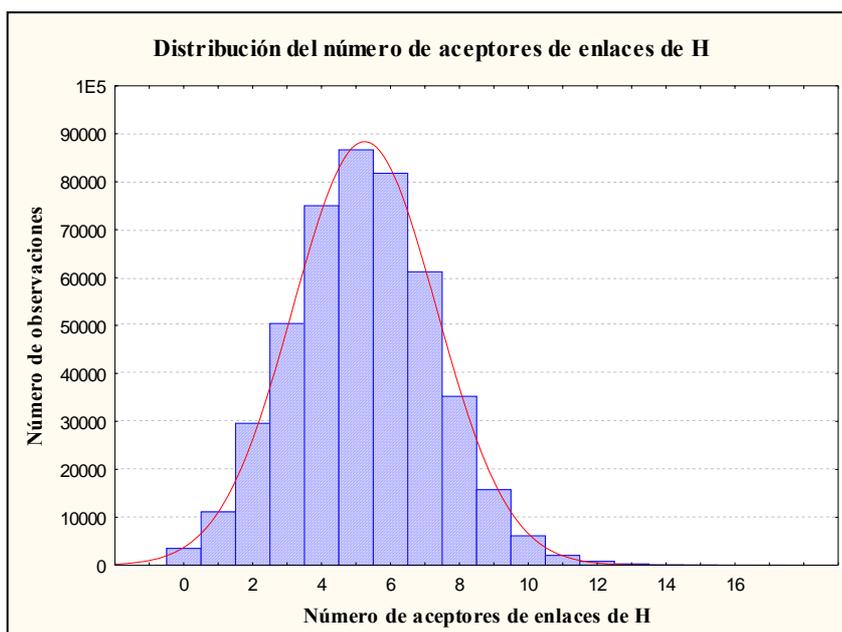


Fig. 8.3. Distribución del número de aceptores de enlaces de H para las 459,718 estructuras de la base de datos.

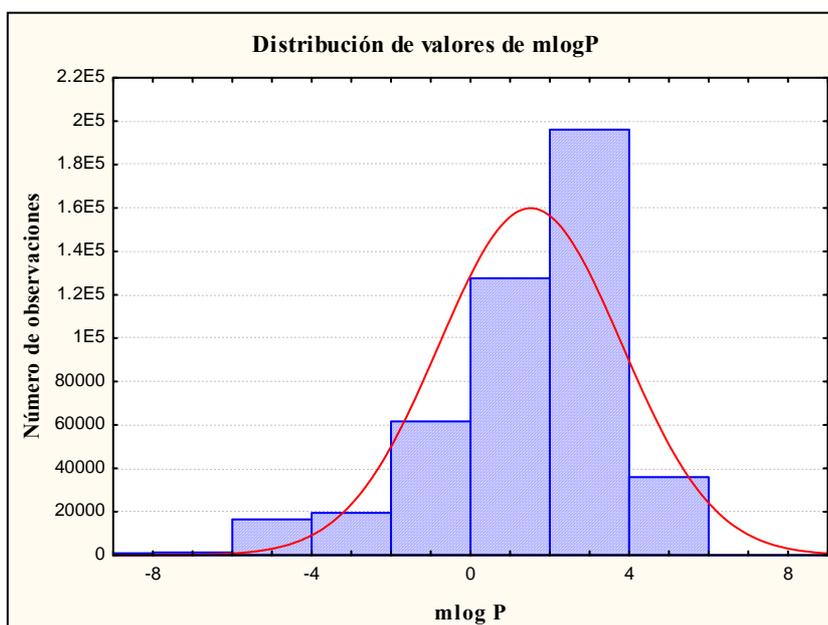


Fig. 8.3. Distribución de valores de mlogP para las 459,718 estructuras de la base de datos.

Por otro lado, en el caso de fármacos anticonvulsivos, como en el caso general de fármacos que actúan a nivel del SNC, el pasaje del



fármaco a través de la BBB es un paso crítico para ejercer la actividad terapéutica. La BBB es una estructura formada por las células endoteliales de los capilares del cerebro y la médula espinal; éstas forman entre sí uniones estrechas debidas a la presencia de proteínas transmembrana que bloquean la libre difusión paracelular de solutos acuosos, restringiendo bidireccionalmente el pasaje de compuestos polares [8]. Se ha determinado que un valor óptimo del logP para difusión pasiva a través de la barrera hematoencefálica de 2 [9,10]. En base a lo expuesto se ha aplicado un filtro secundario consistente en retener, en el proceso de SV, sólo aquellas estructuras con mlog P entre 1.0 y 3.0. Este filtro constituye además una aproximación indirecta para seleccionar estructuras seguras: los compuestos de log P bajo tienden a eliminarse con mayor facilidad (principalmente a través de la orina) mientras que los fármacos con logP elevado tienden a una mayor acumulación en tejidos y deben metabolizarse más extensamente en el hígado para posibilitarse su eliminación a través de la orina, aumentando la probabilidad de generarse de esta manera metabolitos tóxicos [11]. Un ejemplo clásico de esta situación son las vitaminas: las vitaminas liposolubles (vit. A, D, E y K) son tóxicas por bioacumulación si se ingieren por encima de la dosis diaria recomendada, mientras que no ocurre así con las vitaminas hidrosolubles (B y C) que se eliminan fácilmente en orina.

Las 2,719 estructuras seleccionadas mediante SV y aplicación de los filtros generales fueron posteriormente sometidas a los filtros particulares (modelos de permeación a través de la BBB y neurotoxicidad) que se describen en el apartado siguiente.



Filtros particulares.

Se generaron, mediante MLR - Stepwise Forward, filtros para cuantificar la permeabilidad a través de la BBB y la capacidad de producir ataxia en el modelo del Rotorod. En ambos casos se utilizó una tolerancia de 0.5 para evitar la inclusión de pares de descriptores redundantes.

El **modelo de permeación a través de la BBB** se generó a partir de 75 compuestos estructuralmente diversos extraídos de dos estudios similares llevados a cabo por Platts y colaboradores y Pan y colaboradores [12,13]. La variable dependiente modelada fue el log BB, siendo BB la relación entre las concentraciones de equilibrio en cerebro y sangre. Siguiendo la idea de entrenar el modelo con compuestos de características similares a aquellos en los que se va a aplicar, se procuró incluir en el conjunto de entrenamiento moléculas “drug-like”, excluyendo estructuras que los autores anteriores incluyeron en sus respectivos conjuntos de entrenamiento pero que no poseían características estructurales propias de un fármaco (por ejemplo, gases nobles, pentano, hexano, benceno, tolueno, etc.). Las estructuras incluidas en el conjunto de entrenamiento pueden verse en las figs. 8.5a y 8.5b. En la fig. 8.6 se presenta la distribución de valores de log BB para las 75 estructuras del conjunto de entrenamiento; como puede verse, se encuentran bien distribuidos a lo largo de más de cuatro órdenes logarítmicos.

El **modelo de neurotoxicidad** se generó a partir de un conjunto de entrenamiento de 29 estructuras con TD_{50} (ratón, vía de administración intraperitoneal) conocida, determinada mediante el ensayo de rotorod. Los datos de TD_{50} fueron extraídos de literatura [14-22]. Las



estructuras de los compuestos pueden verse en la fig. 8.7. Los valores de TD_{50} se hallan distribuidos a lo largo de sólo dos órdenes logarít-

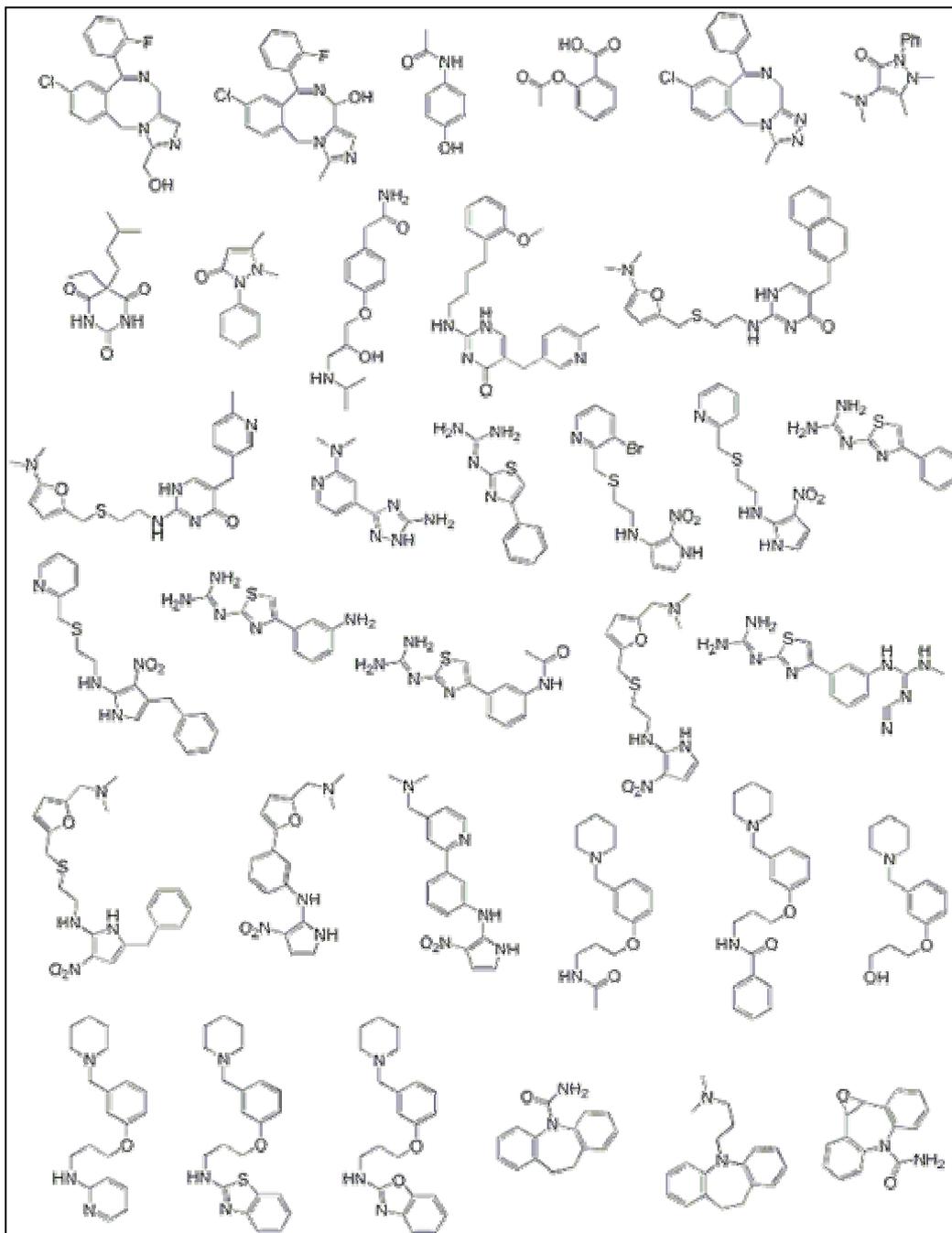


Fig. 8.5a. Estructuras de los compuestos que componen el conjunto de entrenamiento del modelo de permeabilidad (continúa en página siguiente).

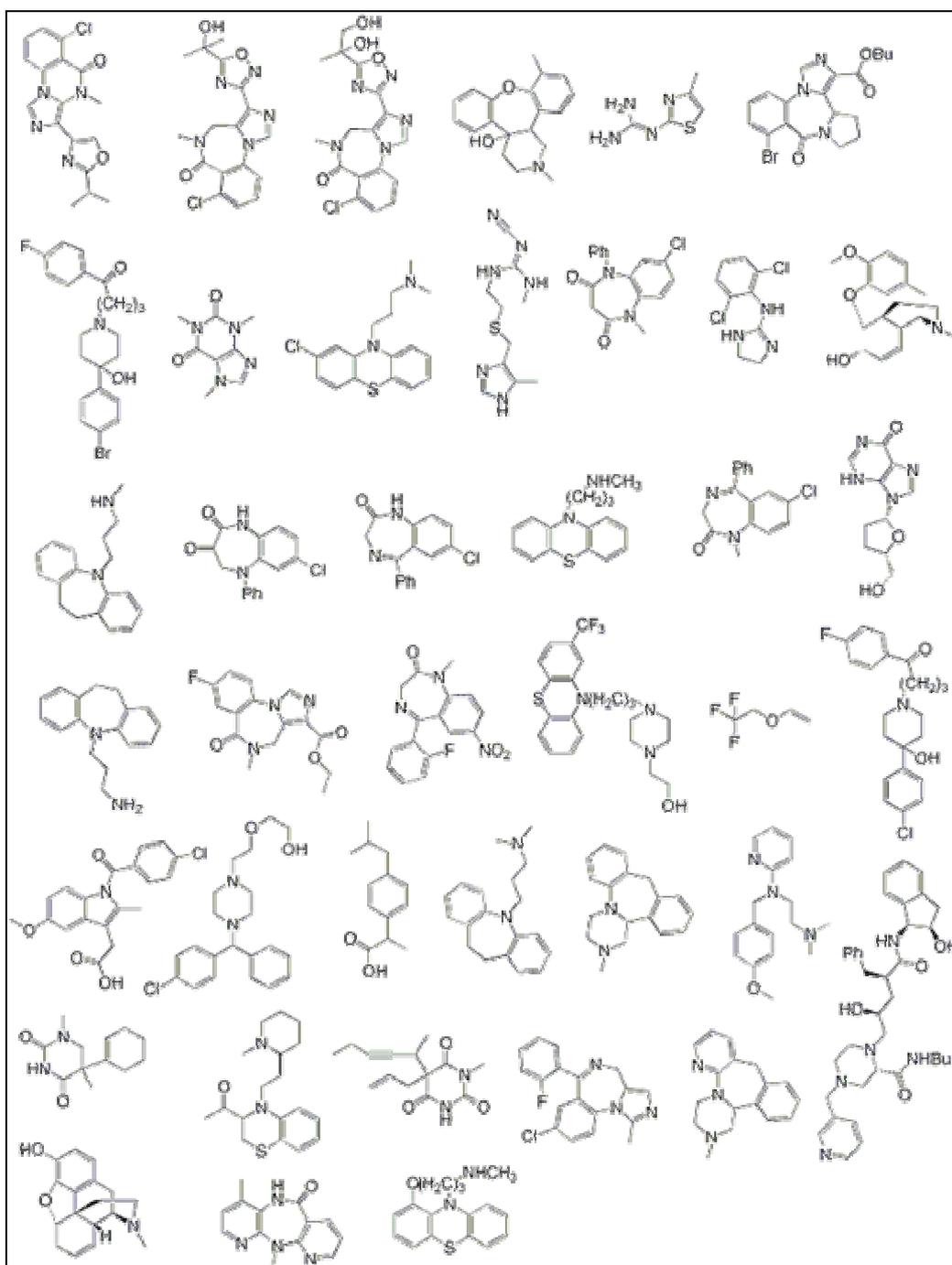


Fig. 8.5b. Estructuras de los compuestos que componen el conjunto de entrenamiento del modelo de permeabilidad (viene de página anterior).

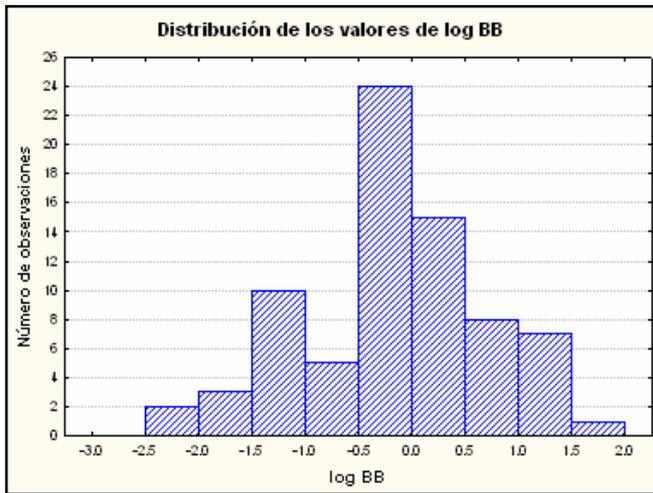


Fig. 8.6. Distribución de valores de log BB para los 75 compuestos que constituyen el conjunto de entrenamiento para el modelo de permeabilidad a través de la barrera hematoencefálica.

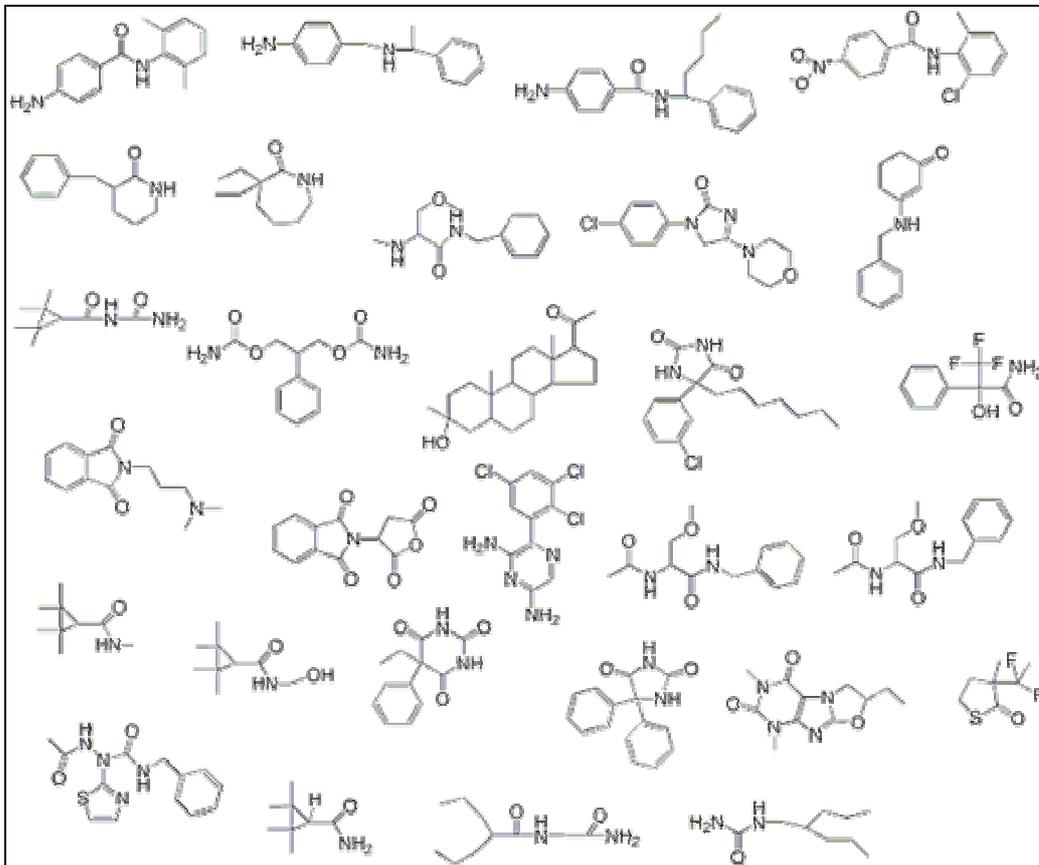


Fig. 8.7. Estructuras de los compuestos que componen el conjunto de entrenamiento del modelo de neurotoxicidad.

micos (log TD_{50} entre 1.6 y 3.6); sin embargo, no se han encontrado datos en literatura fuera de ese intervalo.



Los modelos obtenidos fueron validados mediante validaciones L00 y LG0 (removiendo, sistemáticamente, siete grupos de 10 compuestos y un grupo de 5 en el caso del modelo de permeabilidad, y 4 grupos de 6 compuestos y un grupo de 5 en el caso del modelo de neurotoxicidad), ensayo de aleatorización y validación externa con un conjunto independiente de 40 compuestos (modelo de permeabilidad) y un conjunto independiente de 12 compuestos (modelo de neurotoxicidad).

Los modelos generados y los resultados de la validación L00 y LG0 para los mismos fueron:

- $\log BB = 2.3723 + 0.5041 \times GATS6v + 0.2506 \times nX + 0.2072 \times MLOGP + 0.2009 \times H-053 - 0.1255 \times PHI - 1.1460 \times IC1$

$$N=75 \quad r^2 = 0.76 \quad F = 34.97 \quad SEE = 0.44$$
$$q^2_{L00} = 0.70 \quad SEE_{L00} = 0.49 \quad q^2_{LG0} = 0.62 \quad SEE_{LG0} = 0.53$$

Donde GATS6v es la autocorrelación de Geary de orden 6 ponderada por volúmenes atómicos de Van der Waals; nX representa el número de átomos de halógenos; mlogP es el coeficiente de reparto octanol-agua según Moriguchi; H-053 representa hidrógenos unidos a un átomo de carbono sp³ a) unido a su vez a un átomo de oxígeno y b) adyacente a otro carbono unido a dos heteroátomos; PHI representa el índice de flexibilidad de Kier; IC1 representa, como ya hemos definido, el contenido de información de primer orden. Se observa que, como es bien sabido, la permeabilidad aumenta con el valor del log P; por otro lado, la



presencia de heteroátomos (que tiende a aumentar el valor de IC1 por generar una mayor diversidad en el conjunto de vecindades) disminuye la permeabilidad, aunque en menor medida si los heteroátomos son halógenos (nótese que a mayor valor del descriptor nX mayor la permeabilidad) por cuanto los halógenos no son capaces de formar enlaces de puentes de H siendo que la capacidad de la molécula para establecer puentes de H en general atenta contra la permeabilidad a través de la BBB. Asimismo, la abundancia de pares de átomos formados por átomos de volumen atómico desigual separados a una distancia topológica de 6 aumenta la permeabilidad; esto tiene sentido si se considera que valores altos de GATS6v podrían ser indicativos de una distribución no uniforme de los heteroátomos en la molécula. En contraste, la flexibilidad, probablemente por una cuestión entrópica, atenta contra la absorción.

Puede observarse que el modelo permite predecir valores de log BB con un error cercano media unidad logarítmica. La relación entre el número de casos y el número de descriptores (12.5) es adecuada. Los resultados de ambas validaciones cruzadas son buenos (la SEE de ambas es similar a la del modelo original, y q^2 es mayor a 0.5 en ambos casos). En la validación externa el 85% de los compuestos fueron predichos con un residual menor a 2 SEE.

- $\text{Log TD}_{50} = 1.5758 - 0.3411 \times \text{Hy} - 0.8715 \times \text{O-059} - 0.1423 \times \text{nCIR} + 0.0308 \times (\text{D/D})^{1/3} + 0.6524 \times \text{BELm2}$

N=29 $r^2 = 0.82$ F = 20.38 SEE = 0.22

$q^2_{\text{LOO}} = 0.74$ $\text{SEE}_{\text{LOO}} = 0.26$ $q^2_{\text{LGO}} = 0.66$ $\text{SEE}_{\text{LGO}} = 0.30$



donde H_y representa el factor hidrofílico; 0-059 representa el número de grupos éter; nCIR representa el número de circuitos que pueden trazarse en la molécula; $(D/D)^{1/3}$ es un índice definido por nuestro grupo de trabajo como la raíz cúbica de la relación entre el índice Detour y el número de átomos dadores de enlaces de H en una molécula (que ha demostrado ser inversamente proporcional a la solubilidad acuosa) [23,24] y BELm2 representa el segundo menor autovalor de la matriz Burden ponderada por la masa atómica (esta matriz topológica posee los elementos diagonales iguales a la masa atómica de los átomos; a los elementos no diagonales correspondientes a átomos unidos por un enlace químico se les asigna un valor igual al orden de enlace sobre 10; a los elementos correspondientes a un enlace terminal se les suma 0.01 y todos los demás elementos no diagonales valen 0.001). Si bien el modelo no es fácil de interpretar, puede observarse que a mayor hidrofilia/solubilidad acuosa menor neurotoxicidad (tal vez por una menor llegada al sistema nervioso central y una mayor facilidad de eliminación de sustancias hidrofílicas) y que la neurotoxicidad parece disminuir, dentro de los rangos de descriptores que abarca el conjunto de entrenamiento, con la presencia de grupos éter y de anillos en la estructura (en particular si se trata de ciclos fusionados, ya que de esta manera aumenta el número de circuitos que pueden recorrerse en la molécula).

En este caso las validaciones cruzadas L00 y L01 también dieron buenos resultados; la validación externa arrojó residuales menores a 0.5 unidades logarítmicas en el 75% de los compuestos.

Las figs. 8.8 a 8.11 presentan los resultados del test de aleatorización y los gráficos de residuales versus residuales en



la validación cruzada L00 para ambos modelos, observándose resultados que indican robustez en ambos casos.

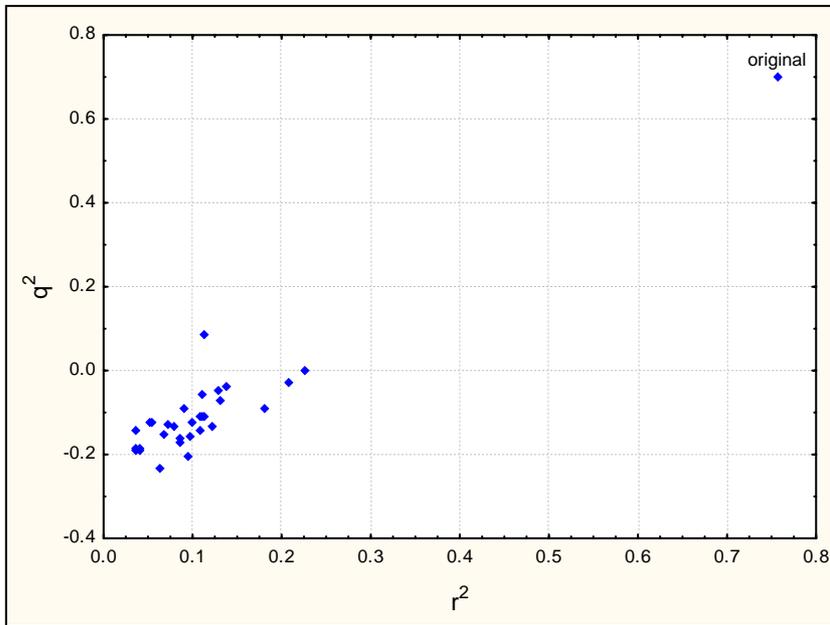


Fig. 8.8. Resultados del ensayo de aleatorización del modelo de permeabilidad. Se observa que el modelo original es estadísticamente superior a los modelos que se obtienen por aleatorización.

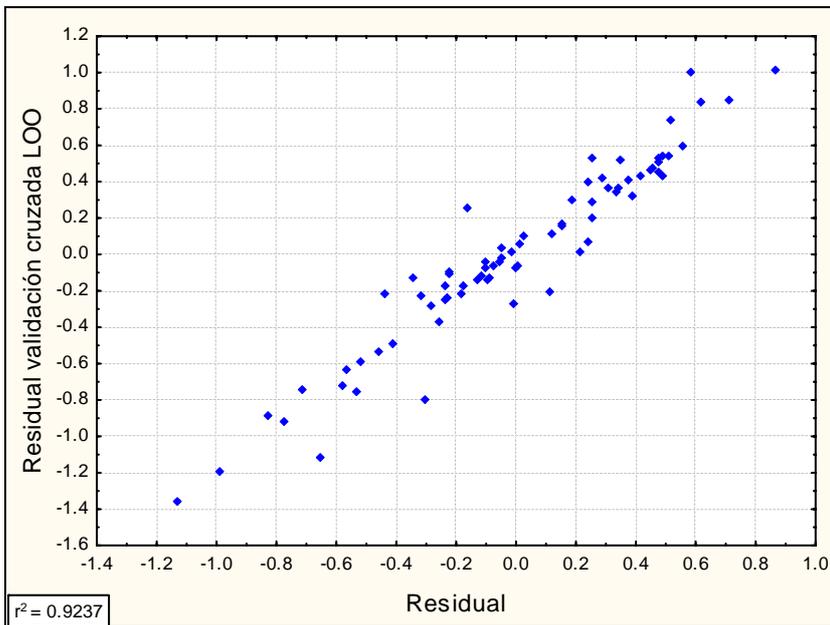


Fig. 8.9. Resultados de la validación cruzada LOO. La buena correlación entre los residuales LOO y los residuales del modelo original indica la robustez del modelo.

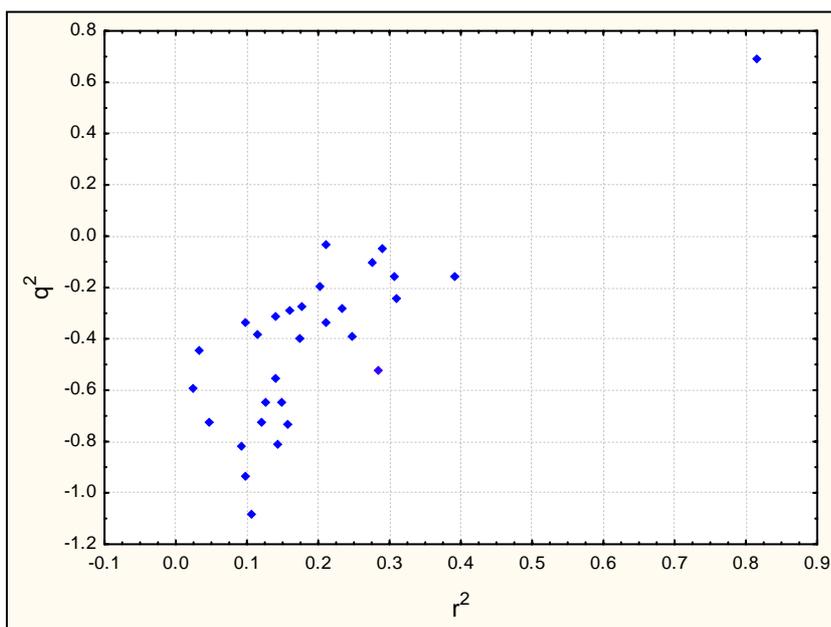


Figura 8.10. Resultados del ensayo de aleatorización para el modelo de neurotoxicidad.

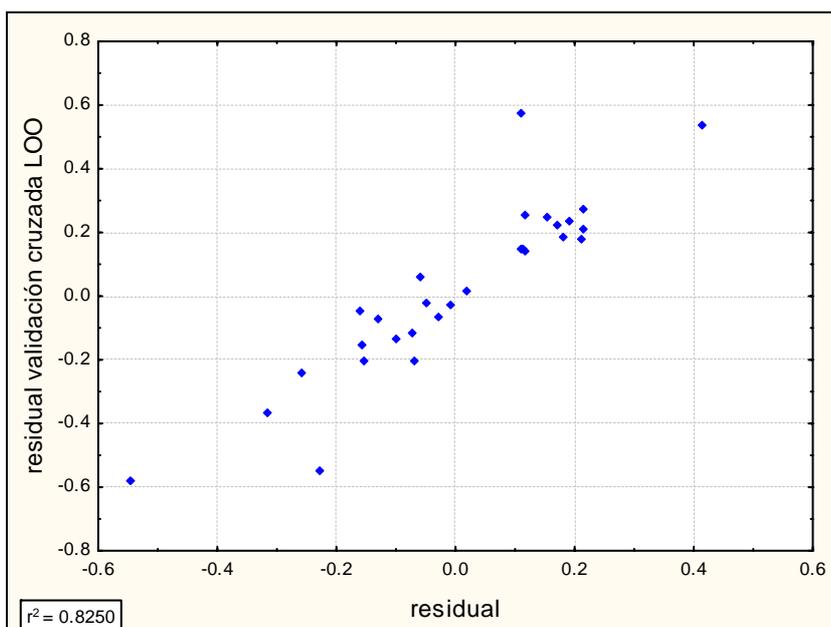


Figura 8.11. Gráfico de residuales LOO versus residuales del modelo original.

De las 2,719 estructuras seleccionadas por la df y los filtros generales a partir de las 459,718 estructuras de la base de datos, se seleccionaron aquellas que, según los modelos anteriores poseen un log BB mayor a -0.5 (indicando que en el equilibrio, la concentración en



cerebro es al menos un 30% de la concentración en sangre) y un TD_{50} en el ensayo de Rotorod mayor a 300 mg/kg (como según el programa del NIH se seleccionarán aquellos fármacos activos a 30 y 100 mg/kg, estaríamos ante un índice terapéutico mínimo cercano a 3). 2,227 de esas estructuras cumplieron con la restricción impuesta por el modelo de permeabilidad, 2,078 superaron el filtro de neurotoxicidad y sólo 1,740 cumplieron ambos requisitos en forma simultánea. En la figura 8.12 se presentan veinte de las estructuras elegidas, con sus respectivos valores de df , $\log BB$ (predicho) y TD_{50} en el ensayo de Rotorod (predicho).

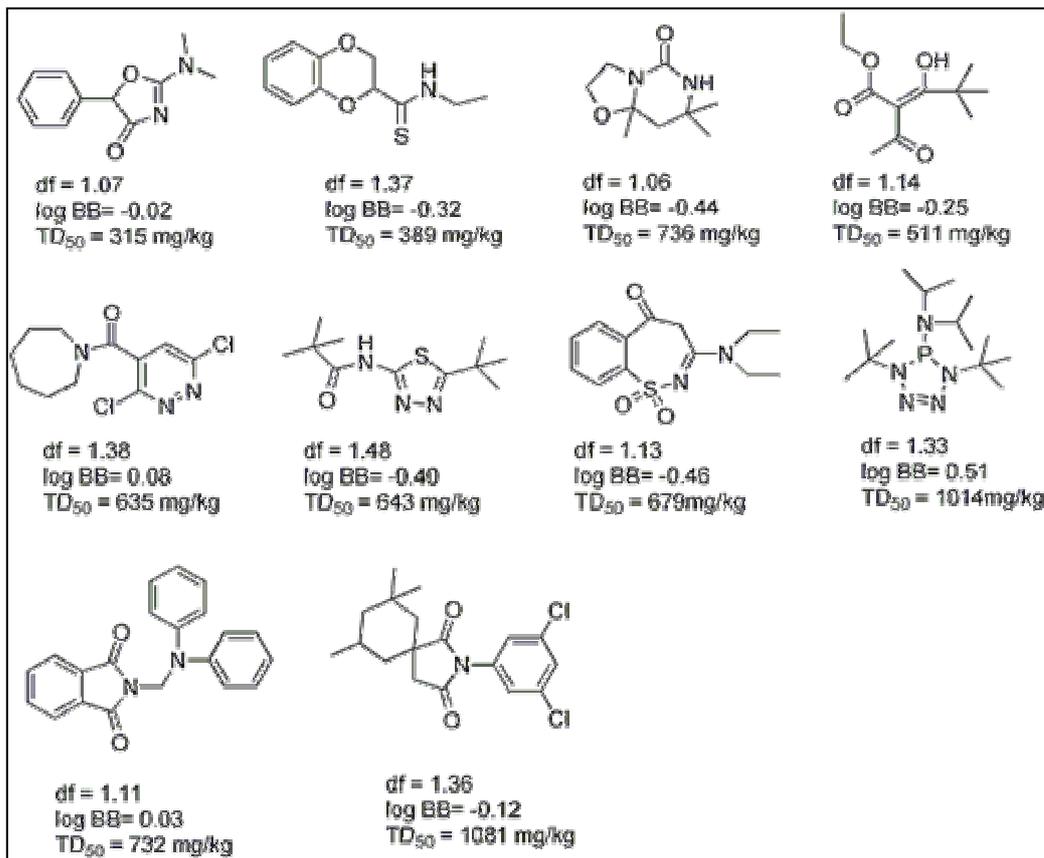


Fig. 8.12. 10 de las estructuras que superaron todos los filtros en el proceso de SV. Se incluyen los valores predichos de df , $\log BB$ y TD_{50} .



De esta manera, mediante aplicación de la *df*, la regla de Lipinski, el valor óptimo del log P para permeación a través de la BBB y los filtros particulares de permeación y neurotoxicidad hemos arribado a 1,740 estructuras potencialmente anticonvulsivas, biodisponibles a nivel del SNC tras administración oral y con bajas probabilidades de producir neurotoxicidad.

Referencias

1. Eddershaw, P.J. ADME/PK as part of a rational approach to drug discovery. *Drug Discov. Today*. **2000**, 5, págs. 409-414
2. Atterwil, C.K.; Wing, M.G. In vitro preclinical lead optimization technologies (PLOTs) in pharmaceutical development. *Toxicol. Lett.* **2002**, 127, págs. 143-151
3. Roberts, S.A. Drug metabolism and pharmacokinetics in drug discovery. *Curr. Opin. Drug. Discov. Dev.* **2003**, 6, págs. 66-80
4. Lin, J.; Sahakian, D.C.; de Morais, S.M.; Xu, J.J.; Polzer, R.J.; Winter The role of absorption, distribution, metabolism and toxicity in drug discovery. *Curr. Top. Med. Chem.* **2003**, 3, págs. 1125-1154
5. Clark, D.E.; Pickett, S.D. Computational methods for the prediction of “drug-likeness”. *Drug. Discov. Today*. **2000**, 5, págs. 49-58
6. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, 46, págs. 3-26
7. Lipinski, C.A. Lead- and drug-like compounds: the rule of five revolution. *Drug Discov. Today: Technol.* **2004**, 1, págs. 337-341
8. Begley, D.J. Delivery of therapeutic agents to the central nervous system: the problems and the possibilities. *Pharmacol. Therapeut.* **2004**, 104, págs. 29-45
9. Ter Laak, A.M.; Tsai, R.S.; Donné-Op den Kelder, G.M.; Carrupt, P.A.; Testa, B.; Timmerman, H. *Eur. J. Pharm. Sci.* **1994**, 2, págs. 373-384
10. Hadjipavlou-Litina, D. Review, reevaluation, and new results in quantitative structure-activity studies of anticonvulsants. *Med. Res. Rev.* **1998**, 18, págs. 91-119
11. Smith, C.J.; Hansch, C. The relative toxicity of compounds in mainstream cigarette smoke condensate. *Food. Chem. Toxicol.* **2000**, 38, págs. 637-646
12. Platts, J.A.; Abraham, M.H.; Zhao, Y.H.; Hersey, A.; Ijaz, L.; Butina, D. Correlation and prediction of a large blood-brain



- distribution dataset - an LFER study. *Eur. J. Med. Chem.* **2001**, 36, págs. 719-730
13. Pan, D.; Manisha, I.; Liu, J.; Li, Y.; Hopfinger, A.J. Constructing optimum blood brain barrier QSAR models using a combination of 4D-molecular similarity measures and cluster analysis. *J. Chem. Inf. Comput. Sci.* **2004**, 44, págs. 2083-2098
 14. Andurkar, S.V.; Béguin, C.; Stables, J.P.; Kohn, H. Synthesis and Structural Studies of Aza Analogues of Functionalized Amino Acids: New Anticonvulsant Agents. *J. Med. Chem.* **2001**, 44, págs. 1475-1478
 15. Luszczkia, J.J.; Czuczvara, S.J. Interaction between lamotrigine and felbamate in the maximal electroshock-induced seizures in mice: an isobolographic analysis. *Eur. Neuropsychopharm.* **2005**, 15, págs. 133-142
 16. Schenck, H.A.; Lenkowski, P.W.; Choudhury-Mukherjee, I.; Ko, S.; Stables, J.P.; Patel, M.K.; Brown, M.L. Design, synthesis and evaluation of novel hydroxyamides as orally available anticonvulsants. *Bioorg. Med. Chem.* **2004**, 12, págs. 979-993
 17. Isoherranen, N.; Levy, R.H.; Yagen, B.; Woodhead, J.H.; White, H.S.; Bialer, M. Metabolism of a new antiepileptic drug, N-methyl-tetramethylcyclopropanecarboxamide, and anticonvulsant activity of its metabolites. *Epilepsy Res.* **2004**, 58, págs. 1-12
 18. Béguin, C.; LeTiran, A.; Stables, J.P.; Voyksner, R.D.; Kohn, H. N-Substituted amino acid NO-benzylamides: synthesis, anticonvulsant, and metabolic activities. *Bioorg. Med. Chem.* **2004**, 12, págs. 3079-3096
 19. Bialer, M.; Johannessen, S.I.; Kupferberg, H.J.; Levy, R.H.; Loiseau, P.; Perucca, E. Progress report on new antiepileptic drugs: a summary of the fourth Eilat conference (EILAT IV). *Epilepsy Res.* **1999**, 34, págs. 1-41
 20. Obniska, J.; Kaminski, K.; Zagorska, A.; Dzierzawska-Majewska, A.; Karolak-Wojciechowska, J. Synthesis and anticonvulsant activity of new fluorinated N-phenyl- and N-benzyl-2-azaspiro[4.4]nonane- and [4.5]decane-1,3-dione derivatives: Part III. *J. Fluorine Chem.* **2006**, 127, págs. 417-425
 21. Li, X.; Morrow, D.; Witkin, J.M. A pharmacophore derived phenytoin analogue with increased affinity for slow inactivated sodium channels exhibits a desired anticonvulsant profile. *Life Sci.* **2006**, 78, págs. 1933-1939
 22. Abdel-Hafez, A.A. Synthesis and Anticonvulsant Evaluation of N-Substituted-Isoindolinedione Derivatives. *Arch Pharm Res.* **2004**, 27, págs 495-501
 23. Talevi, A.; Castro, E.A.; Bruno-Blanch, L.E. New solubility models based on descriptors derived from the detour matrix. *J. Arg. Chem. Soc.* **2006**, 94, págs. 129-141



24. Duchowicz, P.R.; Talevi, A.; Bruno-Blanch, L.E.; Castro, E.A. Application of descriptors based on Lipinski's rules in the QSPR study of aqueous solubility. *Bioorg. Med. Chem.* **2007**, 15, páfs. 3711-3719



Capítulo 8. Otras aplicaciones.

- SV como herramienta de búsqueda racional de reactivos químicos.
- Metodologías integradas de SV 2D y 3D.
- SV por similitud específica.

En el presente capítulo se presentarán brevemente aplicaciones del SV que, si bien no hacen al objetivo principal de esta tesis, han sido desarrolladas como parte de la misma.

SV como herramienta de búsqueda racional de reactivos químicos.

Hasta el momento el SV ha sido ampliamente utilizado en las áreas de la Química y la Fisicoquímica, y en particular en el área de la Química Medicinal, a fin de encontrar sustancias con cierta bioactividad y perfiles toxicológico y farmacocinético adecuados. Durante este trabajo hemos realizado una aplicación original del Análisis Lineal Discriminante: la búsqueda de reactivos químicos específicos para una reacción determinada. En nuestro caso nos ha interesado la búsqueda de nuevos compuestos que sean capaces de transferir (transferidores) el grupo sulfurilo con el fin de ser utilizados en la síntesis de sulfamidas anticonvulsivantes que se realiza en el grupo de trabajo [1,2], particularmente debido a la necesidad de buscar reactivos menos enérgicos y/o selectivos que no reaccionen con otros grupos funcionales presentes en las moléculas de los reactivos. La necesidad de sintetizar sulfamidas anticonvulsivantes se fundamenta en a) la identificación, en el grupo de trabajo, de un farmacóforo antiMES que involucra un grupo amida o carboxilo y que ya ha sido validado experimentalmente con resultados



transferidores del grupo sulfurilo. Para ello se utilizó ALD de manera análoga a la descrita en el capítulo 5. El modelo se generó a partir de un conjunto de entrenamiento de 12 transferidores del grupo sulfurilo obtenidos de literatura (Categoría REACTIVA, Fig. 9.3) [6-9], algunos de los cuales ya han sido utilizados en el laboratorio de Química Medicinal, y 14 sulfamidas estables, sin capacidad aparente de transferencia del grupo sulfurilo (Categoría NO REACTIVA, fig. 9.4).

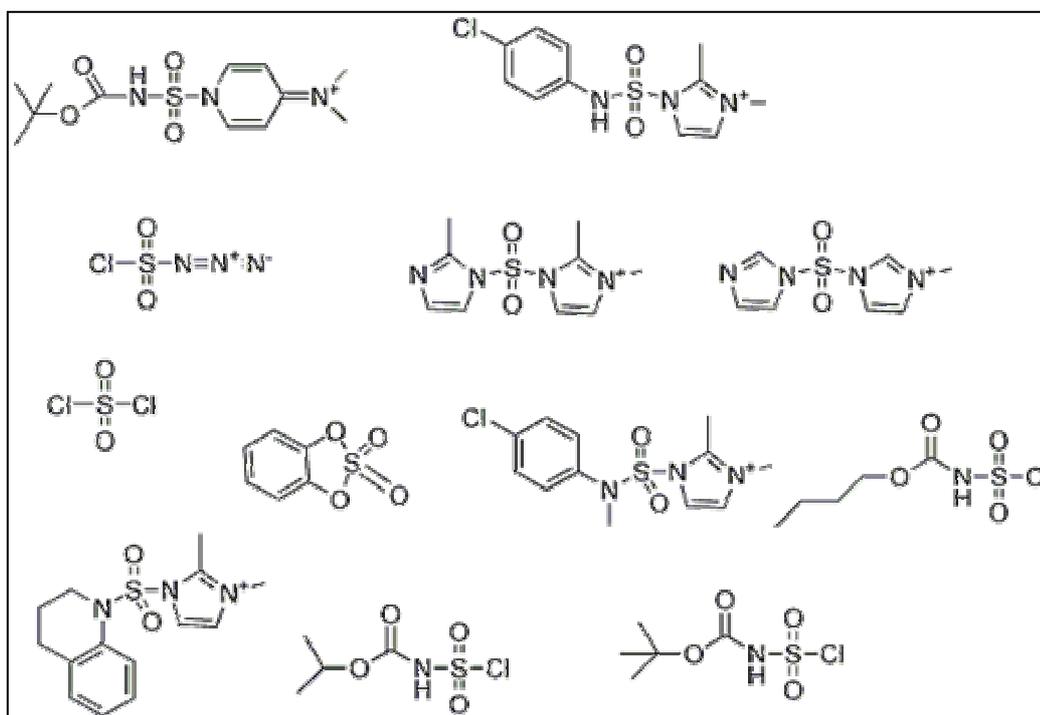


Fig. 9.3. Transferidores conocidos del grupo sulfurilo que conforman la categoría REACTIVA del conjunto de entrenamiento.

El modelo generado obtenido fue:

$$df = 12.5057 - 0.0001 \times \text{GMTIV} + 0.7216 \times \text{C-005} - 1.1717 \times \text{O-058} + 3.3749 \times \text{MAXDN}$$

$$N = 26 \quad \lambda \text{ de Wilks} = 0.37 \quad F = 8.88 \quad p < 0.0002$$



donde GMTIV es el índice topológico molecular de Gutman ponderado por las valencias de los vértices, C-005 es el número de átomos de carbono en la molécula unidos a tres heteroátomos, O-058 es el número de grupos carbonilo y MAXDN es la variación electrotopológica negativa máxima.

El porcentaje de buenas clasificaciones en el conjunto de entrenamiento fue del 92%. El modelo fue validado externamente con un conjunto de prueba de 10 sulfamidas estables. Sólo se incluyeron en el conjunto de prueba compuestos NO REACTIVOS debido al escaso número de compuestos REACTIVOS conocidos (se prefirió no validar externamente con compuestos REACTIVOS para no disminuir el escaso número de compuestos de esta clase del conjunto de entrenamiento). El modelo clasificó correctamente nueve de las diez sulfamidas NO REACTIVAS incluidas en el conjunto de prueba.

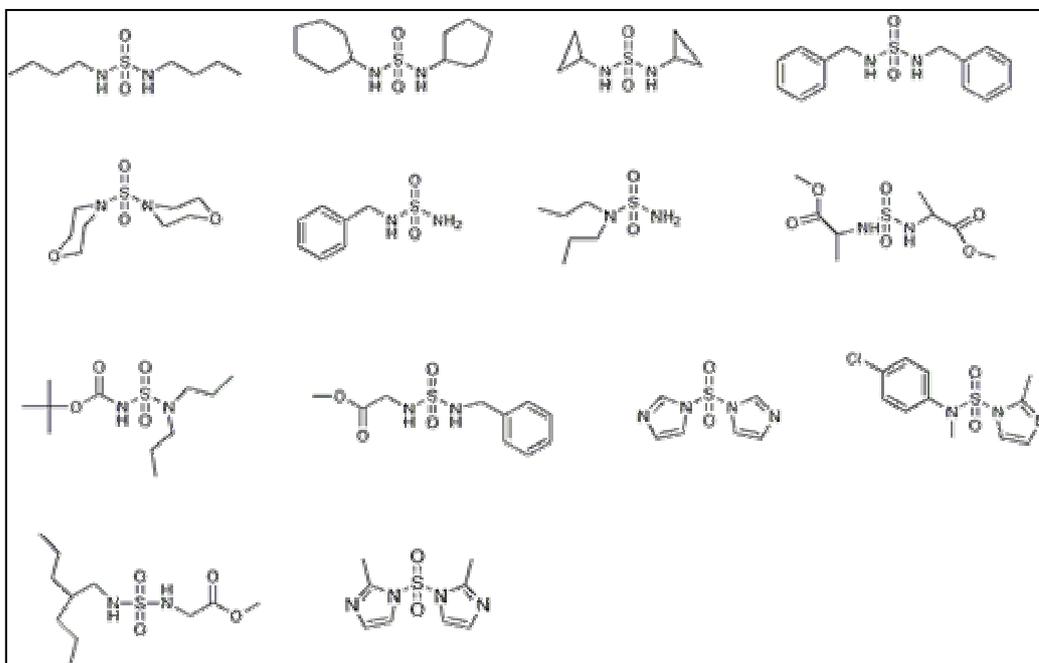


Fig. 9.4. Estructuras que conforman la categoría NO REACTIVA del conjunto de entrenamiento.



Se llevó a cabo el SV de las casi 470,000 estructuras. Se colocó un filtro adicional para excluir aquellos compuestos que no presentaran la subestructura $-SO_2-$. La función discriminante seleccionó

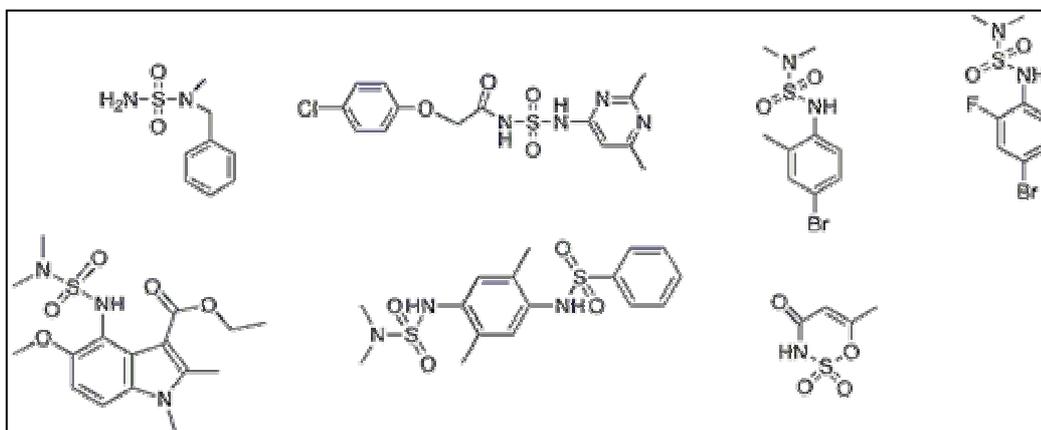
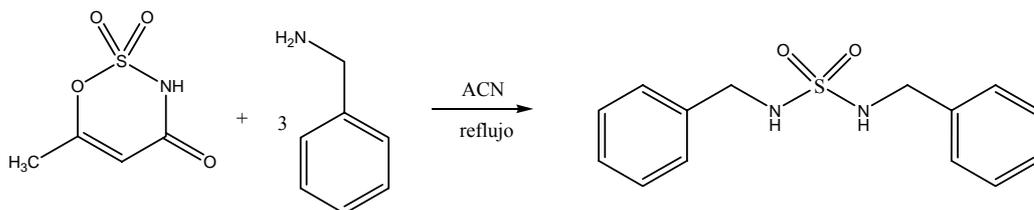


Fig. 9.5. Transferidores del grupo sulfurilo candidatos seleccionados por el SV.

siete potenciales transferidores del grupo sulfurilo, entre ellos, el acesulfame, que ya había sido seleccionado por la *df* para actividad antiMES (ver fig. 9.5).

Se procedió a probar si el acesulfame era, efectivamente, un nuevo transferidor del grupo sulfurilo. Para ello en un tubo de reacción y bajo atmósfera de argón se agregaron 407 mg (2,5 milimoles) de acesulfame y 7 ml de acetonitrilo anhidro (sobre tamiz molecular). Luego se agregaron, gota a gota, 0,7 ml de una solución (669 mg, 6.25 milimoles) de bencilamina (p.eb. 70 C/8mmHg, Δ : 1.68 gr/ml) con 2 ml de acetonitrilo. Se dejó reaccionar durante 26 hs a reflujo. Al cortar el reflujo y alcanzar temperatura ambiente se observó formación de sólido blanco. El solvente fue eliminado en rotavapor. Se secó en tambor de vacío a temperatura ambiente, obteniéndose 942 mg de un sólido blanco amarillento. La reacción propuesta es:



En la fig. 9.6 se presenta el espectro de masas del crudo de reacción; se observa el pico correspondiente al ion molecular de la N,N'-dibencilsulfamida $[M+H]^+ = 277$.

El sólido obtenido se purificó en cromatografía en columna de sílica gel, usando como fase móvil diclorometano:metanol en forma de gradiente desde 80:1 a 10:1. Se recogieron 140 fracciones de 5 ml cada una. La columna se controló mediante TLC, utilizando N,N'-dibencilsulfamida sustancia de referencia. En las fracciones 33 a 53 se observó presencia de un compuesto con Rf igual al de la referencia. Se juntaron dichas fracciones, obteniéndose 30 mg de un sólido blanco. Se realizaron TLCs en diferentes sistemas cromatográficos contra N,N'-dibencilsulfamida sustancia de referencia, observándose Rf coincidente con el de la sustancia de referencia.

Rf del producto en (diclorometano:metanol 40:1): 0.71

Rf del producto en diclorometano: 0.19

Rf del producto en (hexano: acetato de etilo 3:1): 0.35

Por otra parte se juntaron las fracciones 102-130 obteniéndose 730 mg de un producto mayoritario que se encuentra en etapa de purificación.

Los resultados parecen indicar, que en una aplicación original del SV se encontró un nuevo agente transferidor del grupo sulfurilo. Actualmente se están optimizando las condiciones de reacción para favorecer la formación del producto de interés.



De lograr optimizarse la reacción, el acesulfame tendría notables ventajas como transferidor del grupo sulfurilo, dado que es un producto asequible comercialmente y económico, y por sus características es más seguro desde el punto de vista ambiental que otros transferidores utilizados en la actualidad, tales como el cloruro de sulfurilo.

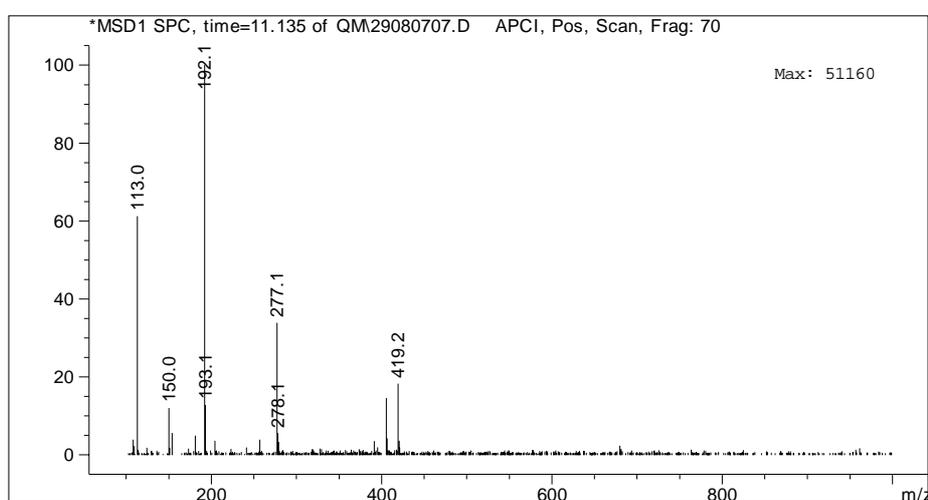


Fig. 9.6. Espectro de masas del crudo de reacción. Se observa presencia de un ión molecular de $[M + H^+] = 277.1$, correspondiente al peso molecular de la N,N'-dibencil sulfamida.

Metodologías integradas de SV 2D y 3D.

Aunque ha habido un extenso debate en literatura respecto a si las metodologías QSAR 3D son mejores que las 2D o viceversa, a lo largo de esta tesis hemos tratado de mantener una posición neutral al respecto, reconociendo que cada metodología presenta ventajas y desventajas y que la elección de una u otra dependerá generalmente de las características del problema que se enfrenta. En el capítulo 4 se ha subrayado que la tendencia moderna es utilizar las distintas aproximaciones en forma complementaria: aprovechar las ventajas de cada una de ellas y compensar las desventajas con el uso



complementario de las restantes. De acuerdo con esta perspectiva, hemos desarrollado un protocolo para la búsqueda de nuevos antiepilépticos que consiste simplemente en el uso secuencial de la df desarrollada y los filtros ADME/Tox generales, en primer lugar, y la superposición con el farmacóforo presentado al inicio de este capítulo, en segunda instancia, calculando mediante métodos cuánticos la diferencia de energía entre la conformación de mínima energía identificada mediante análisis conformacional y la conformación activa definida por el farmacóforo.

En la fig. 9.7 se presenta un esquema de la metodología empleada. Puede observarse que, además de los filtros dados por las condiciones de Lipinski se han aplicado otros filtros de absorción generales: las condiciones de Veber. Las mismas establecen que, para ser biodisponible por vía oral, un compuesto debe tener menos de diez enlaces químicos con libre rotación y un área de la superficie polar menor a 140 \AA^2 [10]. La metodología se ha aplicado a una base de datos de 10,903 productos naturales [11]. Los productos naturales persisten como una de las fuentes más importantes de estructuras líderes de la Química Medicinal. Cerca de un 30% de las casi 900 nuevas entidades químicas introducidas en el mercado entre 1981 y 2002 son productos naturales o derivados hemisintéticos de los mismos, sin tener en cuenta péptidos o proteínas derivados de aplicaciones biotecnológicas [12]. La Química Medicinal moderna busca conseguir gran diversidad estructural dentro de los límites impuestos por el concepto de “drug likeness”. En este sentido los productos naturales son sumamente interesantes, por cuanto presentan una gran diversidad estructural y por sus características estructurales particulares respecto a los productos sintéticos (mayor rigidez estructural, abundancia de átomos de oxígeno) suelen ser lo que se conoce como “tight-binders”, es decir, presentan una potencia relativamente alta [13]; la proporción



de fármacos obtenidos de productos naturales que rompe con más de una de las condiciones de Lipinski es, adicionalmente, similar a la que presentan los fármacos sintéticos. Sin embargo, la investigación de productos naturales por parte de las compañías farmacéuticas ha decaído en las últimas décadas. La razón de este declinar se relaciona con el auge de las metodologías de screening de alta performance: los productos naturales, obtenidos en general luego de realizar uno o varios extractos seguidos de aislamiento bioguiado, elucidación de la estructura y desarrollo de un método para obtener el o los principios activos a escala industrial, son menos compatibles con la velocidad de las tecnologías de HTS (y menos compatibles también con el paradigma de velocidad de la era moderna) que los productos sintéticos [13,14]. Por todo esto, el SV es una metodología atractiva para encarar el screening de grandes colecciones de productos naturales y explotar la inmensa diversidad que éstos ofrecen [14,15].

56 de los 10,903 compuestos fueron seleccionados como candidatos por la df presentada en el capítulo 5 y los filtros generales de absorción ya descritos. Obsérvese que la proporción de compuestos promisorios en el base de datos naturales es casi idéntica a la que se observó en la base general de 459,718 compuestos: alrededor del 0.05%. En la tabla 9.1. se presentan los valores de los descriptores considerados en los filtros generales de absorción. Estos datos parecerían confirmar que los productos naturales son por lo menos tan buenos candidatos como los productos sintéticos, y que la mayor proporción relativa de productos sintéticos introducidos al mercado recientemente se relaciona más con las dificultades ya mencionadas para la identificación del principio activo en una mezcla compleja como son los extractos de productos naturales, su aislamiento y *scaling up*. De nuevo, el SV puede ser una herramienta valiosa para resolver la primera de estas cuestiones. Los compuestos seleccionados



pueden observarse en la figuras 9.8a y 9.8b. Estos compuestos fueron superpuestos con el farmacóforo mediante la herramienta *atom fit* provista por el software Sybyl [16], descartando los que a primera vista no cumplían con los requisitos farmacofóricos. Se tomó el valor de RMS como criterio de la bondad de la superposición:

$$RMS = \sqrt{\frac{\sum d^2}{n}}$$

donde d es la distancia, en Å, entre los pares de átomos superpuestos y n es el número de átomos considerados en la superposición. Para ello se utilizó un criterio conservador: sólo pasaron a la siguiente etapa del protocolo las estructuras con RMS menor o igual a 0.2. En la Tabla 9.2. se presentan los valores de RMS para las estructuras superpuestas, indicando en qué casos no se realizó superposición por no cumplir con los requisitos farmacofóricos. La no superposición con el farmacóforo no implica que esas estructuras no sean activas: como ya se describió, existen diferentes subtipos de canales de sodio con afinidad distinta a las drogas conocidas, y un mismo subtipo de canal de sodio incluye múltiples sitios de unión, por lo que es difícil hallar un farmacóforo único que identifique fármacos que interactúen con todos los sitios activos posibles de los múltiples canales de sodio voltaje operados (más aún: no todas las drogas antiMES evaluadas en este trabajo cumplen con el farmacóforo; tal es el caso de los parabenos y el acesulfame).

Ocho estructuras superaron la etapa de superposición con el farmacóforo. La superposición de las mismas se presenta en la fig. 9.9. Se procedió luego al análisis conformacional sistemático de cada una de ellas para identificar el mínimo global, variando en 30° (hasta

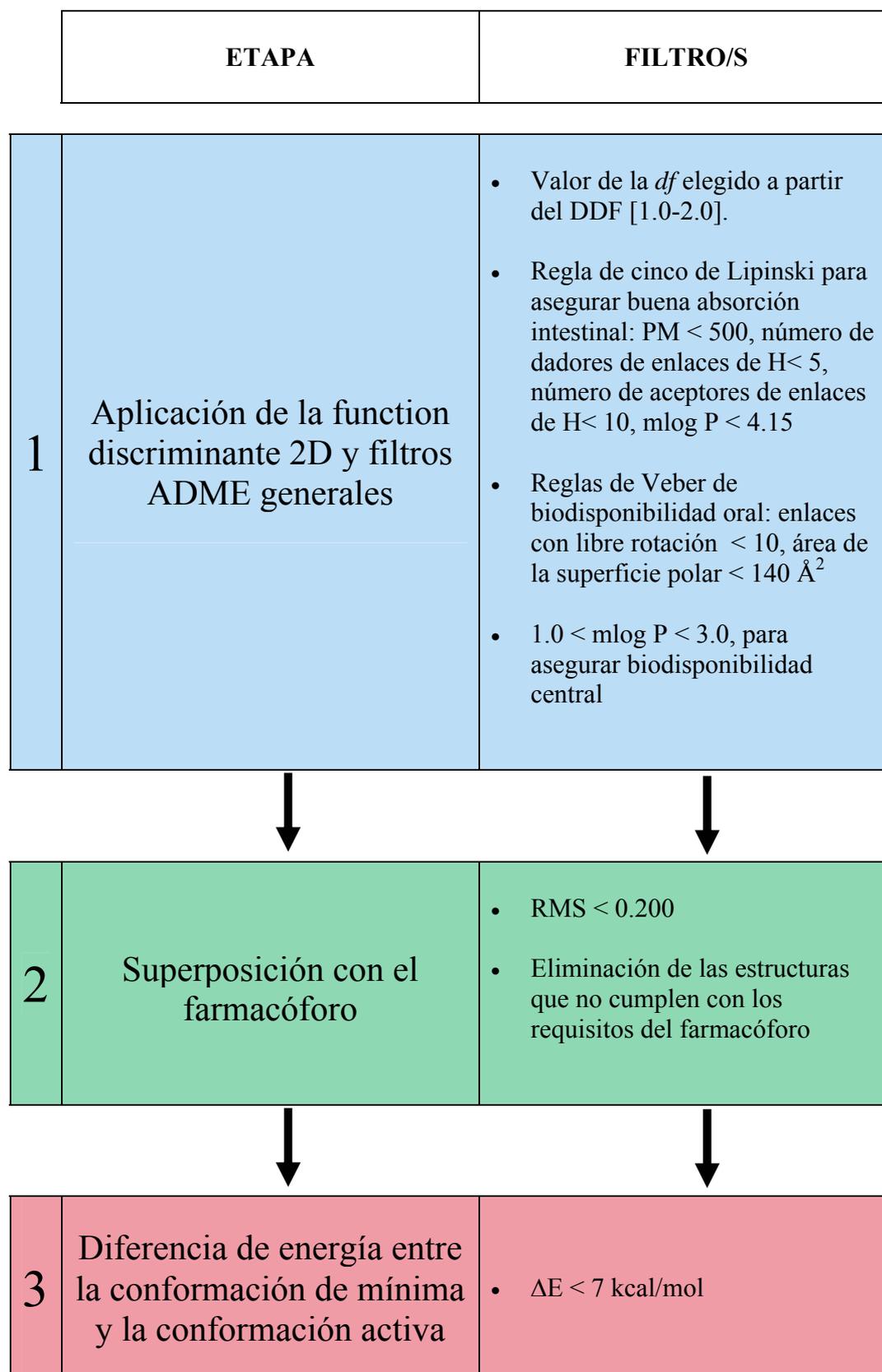


Fig. 9.7. Esquema del protocolo utilizado en la exploración de la base de datos de productos naturales.



completar los 360°) cada uno los ángulos de torsión con libre rotación, y procediendo a la optimización de la geometría a partir de cada una de las geometrías iniciales obtenidas de esa manera. Una vez identificado el mínimo global se procedió a la optimización restringida de geometría, es decir, manteniendo fijos los ángulos correspondientes al farmacóforo en la conformación que consideramos activa (la conformación del farmacóforo). Finalmente, se calculó la diferencia de energía entre la conformación de mínima y la conformación activa definida por el farmacóforo, y se retuvieron las estructuras con una diferencia de energía menor a 7 kcal/mol.

Tabla 9.1. Valores de *df* y valores de los descriptores utilizados en los filtros generales de absorción para las 56 estructuras seleccionadas en la primera etapa del SV. Todos los compuestos presentan biodisponibilidad sistémica según las condiciones de Lipinski y Veber y biodisponibilidad a nivel del SNC según el valor óptimo del coeficiente del logP para la difusión pasiva a través de la BBB. PM: peso molecular; nHDon y nHAc: número de dadores y aceptores de enlaces de H, respectivamente; mlogP: log P calculado según Moriguchi; RBN: número de enlaces con libre rotación; PSA: área de la superficie polar basada en fragmentos.

Compuesto	<i>df</i>	PM	nHDon	nHAc	mlogP	RBN	PSA
1	1.05	217.24	1	4	1.697	4	43.37
2	1.77	180.27	0	2	1.956	6	17.07
3	1.47	287.1	0	4	1.942	4	52.6
4	1.01	248.3	1	4	1.994	2	35.53
5	1.08	262.38	0	3	1.96	7	43.37
6	1.21	220.29	0	3	1.995	6	27.69
7	1.34	205.28	1	3	1.717	6	34.14
8	1.72	251.68	0	4	1.830	4	37.38
9	1.08	245.3	0	4	1.640	3	38.77
10	1.05	278.38	1	4	2.232	7	35.53
11	1.36	228.32	0	3	1.368	4	25.89
12	1.41	225.21	0	4	1.345	0	42.68
13	1.26	278.38	1	4	2.232	7	35.53
14	1.00	198.24	3	3	1.286	2	37.38
15	1.72	219.26	2	4	1.105	3	43.37
16	1.08	240.33	0	3	1.542	5	25.89
17	1.28	213.31	0	3	1.779	2	20.01
18	1.30	204.19	0	4	1.822	3	52.6
19	1.59	206.32	2	3	1.904	5	32.67
20	1.72	202.28	2	3	1.717	5	21.48
21	1.20	242.3	4	5	1.540	5	43.37



22	1.26	228.24	0	4	1.345	2	39.44
23	1.31	205.28	1	3	1.717	7	34.14
24	1.69	230.23	0	4	1.286	3	48.67
25	1.73	377.97	0	5	1.154	4	17.07
26	1.34	183.22	1	2	2.390	1	15.79
27	1.66	229.25	0	4	1.156	4	43.96
28	1.28	264.36	2	5	1.965	8	58.97
29	1.20	234.35	0	3	2.258	7	18.46
30	1.38	274.34	0	3	2.981	1	24.72
31	1.43	244.26	0	4	1.826	2	52.6
32	1.85	228.26	0	3	2.634	3	39.44
33	1.58	271.4	1	4	2.999	10	41.79
34	1.12	262.38	1	3	2.772	7	26.3
35	1.07	226.24	0	3	2.600	3	43.37
36	1.29	214.23	0	3	2.370	2	39.44
37	1.59	242.29	0	3	2.890	4	39.44
38	1.02	228.26	0	3	2.634	3	39.44
39	1.07	302.35	3	4	2.738	3	54.45
40	1.09	203.21	0	4	2.228	3	46.61
41	1.08	199.22	3	3	2.143	3	29.43
42	1.74	193.22	0	4	1.469	5	46.61
43	1.02	232.25	0	4	2.374	5	52.6
44	1.14	238.21	1	6	1.734	8	69.67
45	1.02	263.26	0	4	2.505	0	51.8
46	1.47	253.32	0	3	2.654	5	22.87
47	1.09	203.21	2	4	1.286	3	39.44
48	1.14	240.28	0	5	1.038	5	69.67
49	1.29	150.14	0	3	1.074	1	26.3
50	1.18	204.24	1	3	1.913	5	30.21
51	1.24	248.25	0	4	2.385	0	50.91
52	1.12	218.65	2	3	1.692	0	32.86
53	1.55	180.22	1	3	2.086	6	26.3
54	1.04	282.37	5	4	2.967	5	41.79
55	1.19	210.62	0	3	1.9	2	26.3
56	1.37	242.29	0	3	2.89	4	39.44

Cinco de las estructuras cumplieron con este último requisito, seleccionándose como los mejores candidatos entre las 10,903 estructuras sometidas a la metodología mixta de SV (Tabla 9.3).

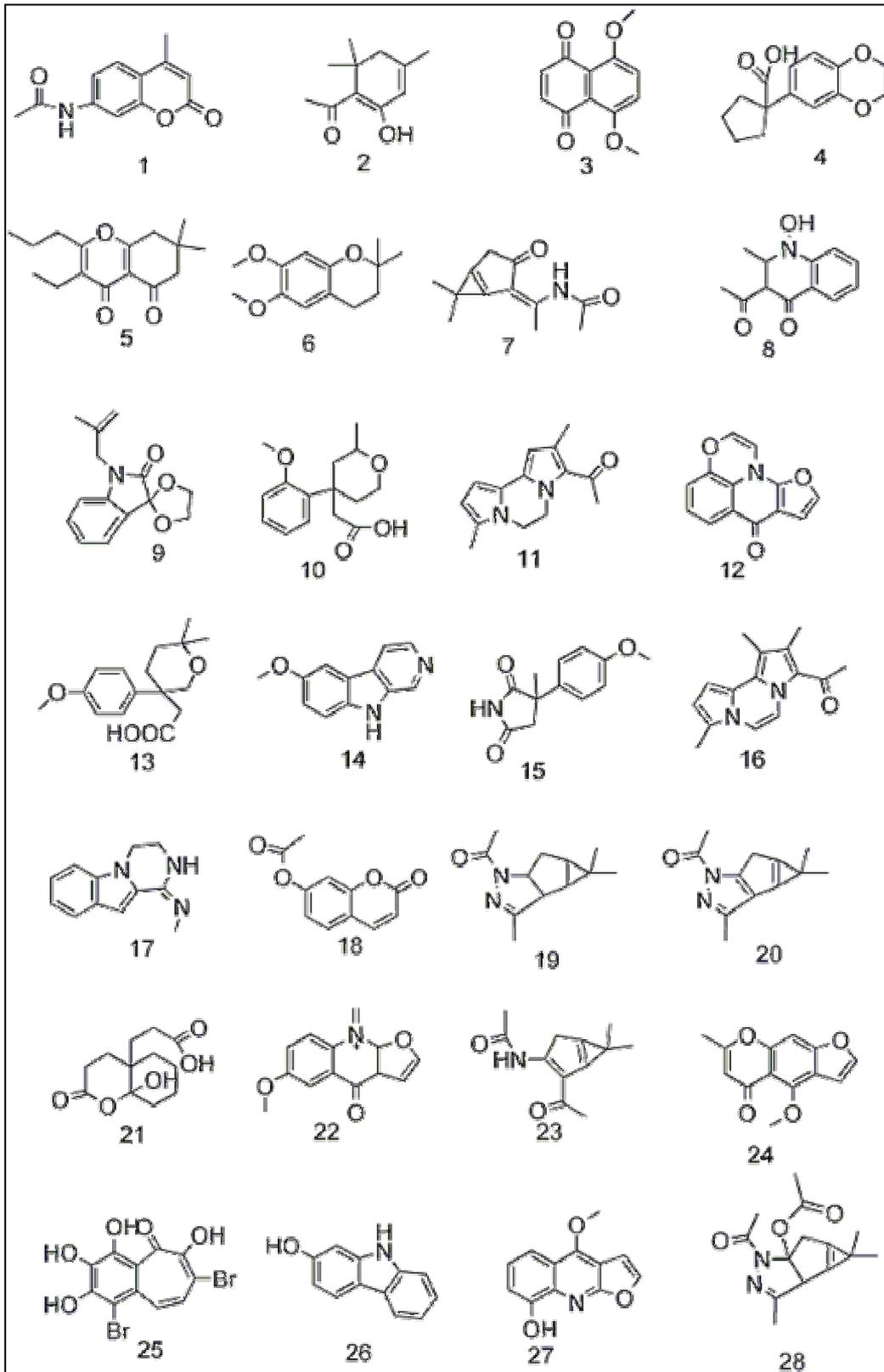


Fig. 9.8a. Estructuras seleccionadas por la *df* y los filtros ADME generales.

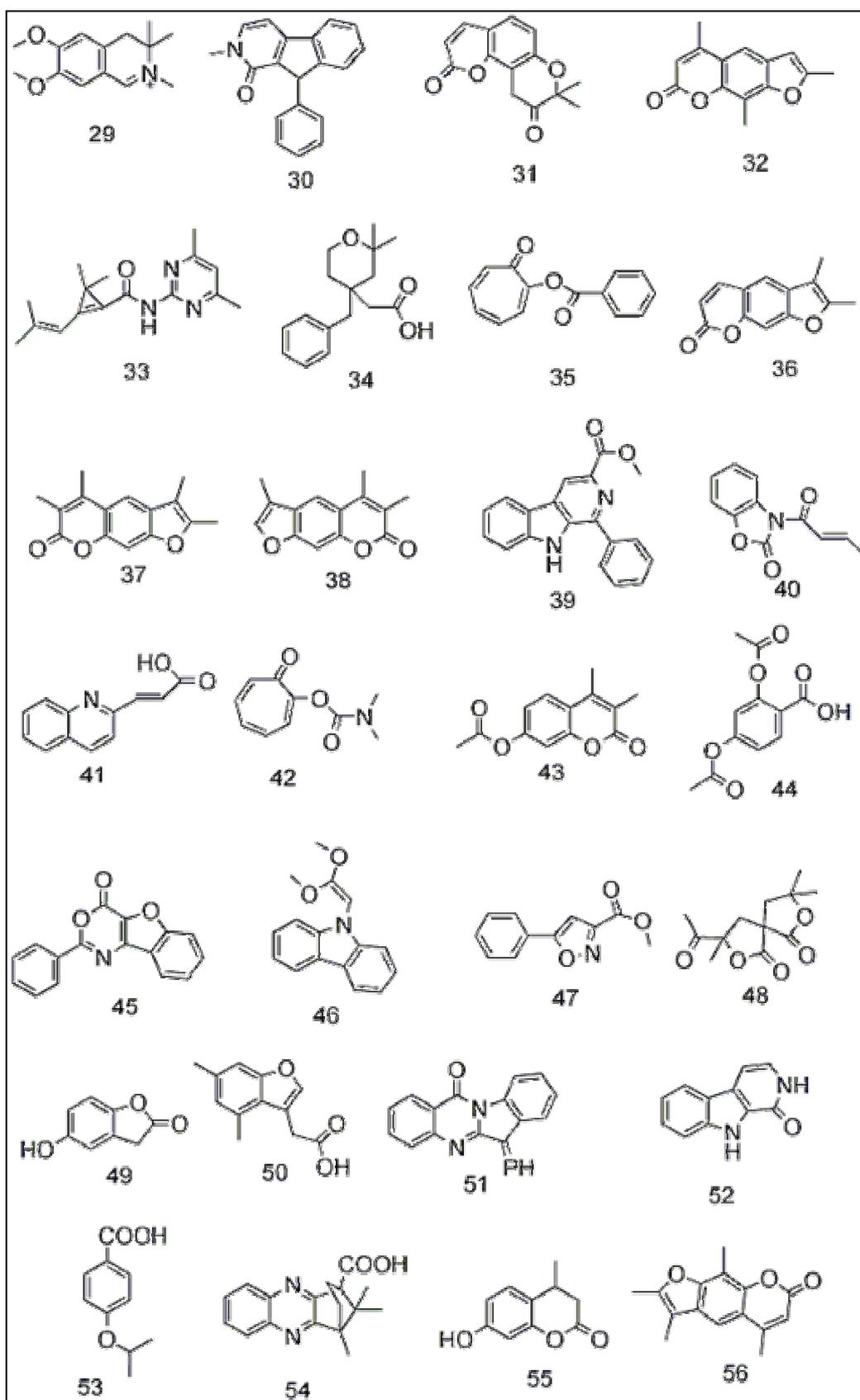


Fig. 9.8b. Estructuras seleccionadas por la *df* y los filtros ADME generales.



Tabla 9.2. Valores de RMS resultantes de la superposición con el farmacóforo de las 56 estructuras seleccionadas en la primera etapa. Se indican en azul las 8 estructuras que pasan a la etapa 3.

Compuesto	RMS	Compuesto	RMS
1	0.146	29	no cumple con el farmacóforo
2	0.670	30	0.418
3	no cumple con farmacóforo	31	0.320
4	no cumple con farmacóforo	32	no cumple con farmacóforo
5	0.472	33	0.182
6	no cumple con farmacóforo	34	0.369
7	no cumple con farmacóforo	35	0.226
8	0.186	36	no cumple con farmacóforo
9	no cumple con farmacóforo	37	no cumple con farmacóforo
10	0.145	38	no cumple con farmacóforo
11	0.853	39	0.227
12	no cumple con farmacóforo	40	0.386
13	0.146	41	0.155
14	0.610	42	0.402
15	0.490	43	no cumple con farmacóforo
16	0.433	44	0.665
17	no cumple con farmacóforo	45	0.368
18	no cumple con farmacóforo	46	no cumple con farmacóforo
19	no cumple con farmacóforo	47	0.320
20	no cumple con farmacóforo	48	no cumple con farmacóforo
21	0.165	49	no cumple con farmacóforo
22	no cumple con farmacóforo	50	0.186
23	no cumple con farmacóforo	51	0.369
24	0.430	52	0.254
25	no cumple con farmacóforo	53	0.697
26	no cumple con farmacóforo	54	0.130
27	0.396	55	no cumple con farmacóforo
28	0.316	56	no cumple con farmacóforo

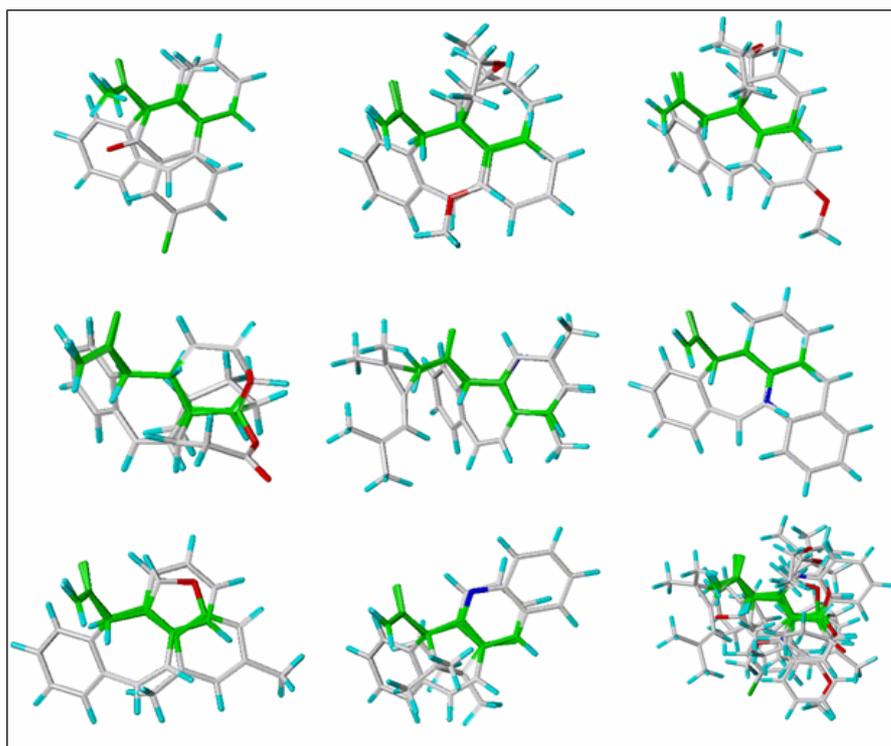


Fig. 9.9. Superposición de las 8 moléculas con $RMS < 0.2$ y el farmacóforo (en verde, los átomos superpuestos). En la esquina inferior derecha se muestra la superposición del conjunto de las ocho estructuras y el farmacóforo.

Tabla 9.3. Diferencias de energía entre la conformación de mínima energía y la conformación activa para las 8 estructuras que llegaron a la etapa 3 del protocolo.

Compuesto	Energía de la conformación de mínima (Hartree)	Energía optimización restringida (Hartree)	AE (Hartree)	ΔE (kcal/mol)
8	-1203.889533	-1203.888290	0.0012425	0.78
10	-923.8746053	-923.8541075	0.0204978	12.86
13	-923.8759499	-923.8611682	0.0147817	9.28
21	-844.2857680	-844.2820324	0.0037356	2.34
33	-861.8181729	-861.8101584	0.0080145	5.03
41	-667.9262359	-667.9131722	0.0130637	8.20
50	-690.2123885	-690.2106883	0.0017002	1.07
54	-918.6504764	-918.6467934	0.0036830	2.31



SV por similaridad específica.

En una revisión reciente sobre el uso de similaridad en SV, Willett señala que, si bien una medida de similaridad involucra tres componentes:

- a) la representación que se usa para caracterizar a las estructuras comparadas (por ej. pares de átomos);
- b) el esquema según el cual se ponderan los componentes de esas representaciones (es decir, ¿se asigna un peso uniforme a todos los componentes o se le da más importancia a algunos que otros? y;
- c) el coeficiente de similaridad que se usa para la comparación

hasta el momento casi no se han realizado estudios respecto a la forma en que se ponderan los componentes de la representación [Willett, op. cit. capítulo 4].

Atendiendo a esta falta de estudios al respecto, y aprovechando que el farmacóforo nos brinda criterios para definir a qué componentes de la representación asignarles mayor importancia, hemos desarrollado un modelo de similaridad específico para la identificación de anticonvulsivantes bloqueantes de canales de sodio. El mismo se basa en la definición de pares de átomos propuesta por Carhart [18]:

(descripción del átomo 1)-(distancia)-(descripción del átomo 2)

donde la descripción del átomo 1 incluye la naturaleza del átomo, su hibridación y el número de átomos distintos de H a él unidos y “distancia” se refiere a la distancia topológica más uno (es decir, el número total de átomos incluidos entre el AP considerado, por el camino más corto). En general, un punto o asterisco al lado del



símbolo atómico indica la presencia de un electrón π y X_n indicará el número n de átomos distintos de H unidos al átomo 1 o 2 mediante enlace covalente.

Para una estructura dada, el número total de pares de átomos que pueden derivarse está dado por $n \times (n-1)/2$, donde n es el número de átomos distintos de H que posee la molécula.

Se propone un algoritmo basado en el coeficiente de similaridad de Tanimoto para identificar nuevos compuestos bioactivos cuando se dispone de un farmacóforo identificado para la actividad de interés:

$$A = S^b \left[1 - \frac{\sum_i \frac{n_i \cdot \Delta n_i}{1 + ABS(\lambda_i - \alpha)}}{c} \right]$$

donde n_i es el número de veces que el AP i aparece en la estructura de referencia, Δn_i es la diferencia entre el número de veces que el AP i aparece en el compuesto que se está evaluando y el número de veces que ese tipo de AP aparece en la referencia, λ_i es la longitud del AP i y α es la máxima longitud de los pares de átomos conteniendo heteroátomos que pueden derivarse del farmacóforo identificado (en nuestro caso, 6). b y c son constantes que moderan la influencia sobre A de S y del segundo término entre paréntesis. Sólo se consideran en el corchete los APs que contengan heteroátomos, dado que los heteroátomos son críticos para el reconocimiento del fármaco por el receptor. S es el coeficiente de Tanimoto en su forma binaria ya definida o en su forma algebraica. Si se comparan dos moléculas A y B , el coeficiente de Tanimoto en su forma algebraica sería:



$$S_{A,B} = \frac{\sum_i na_i \times nb_i}{\sum_i na_i^2 + \sum_i nb_i^2 - \sum_i na_i \times nb_i}$$

donde na_i es el número de veces que el AP del tipo i aparece en la molécula A y nb_i es el número de veces que el AP i aparece en la molécula B. Obsérvese que los tipos de pares de átomos presentes en sólo uno de los dos compuestos comparados no cuentan en el numerador (ya que uno de los factores del producto $na_i \times nb_i$ del término correspondiente de la sumatoria es igual a 0) pero sí en el denominador a través del cuadrado del número de veces que aparece ese par en A o B, según sea el caso.

El valor de cada término i en la sumatoria entre corchetes puede ser interpretado como la contribución del AP i a la actividad de la estructura. El signo de esos términos está dado únicamente por Δn_i , que es el único factor que puede tomar valores mayores o menores que cero. Como el compuesto de referencia será un compuesto muy activo, los APs presentes en el mismo pueden considerarse características “deseadas” involucradas en la interacción con el blanco molecular. Esta es una de las hipótesis fundamentales del modelo propuesto. Si un tipo de AP está presente un número mayor de veces en la referencia que en el compuesto evaluado, entonces el compuesto evaluado tiene una carencia relativa de una característica que favorece la actividad; caso contrario el compuesto evaluado tiene una abundancia relativa de una característica deseada y mayores probabilidades de que la misma se exprese al interactuar con el receptor.

Ahora bien, ¿son todos los tipos de AP igualmente significativos para la actividad? Como de nuevo, la referencia es una estructura muy activa, cuanta mayor cantidad de veces un AP aparezca en la referencia



mayor la probabilidad de que esa subestructura en particular contribuya a la actividad biológica; esto se considera incluyendo n_i en el numerador de los términos de la sumatoria entre corchetes. En el denominador del mismo, adicionalmente, se le da mayor importancia a los APs que presentan una separación igual a la que separación máxima de los APs conteniendo heteroátomos que pueden derivarse del farmacóforo.

Para evaluar la capacidad para identificar fármacos antiMES del modelo se construyeron dos conjuntos de estructuras conteniendo estructuras activas en el ensayo MES y estructuras con otras bioactividades. El conjunto A está formado por 12 anticonvulsivos con actividad alta y moderada en el MES test, estructuralmente diversos, y 11 compuestos con otras bioactividades. También se incluyeron 2 anticonvulsivantes (etosuximida y TV-1901) con baja actividad antiMES. Carbamacepina ($ED_{50} = 37 \mu\text{mol/kg}$) fue definido como el compuesto de referencia del conjunto A. El conjunto B (fig. 9.11) incluye 26 estructuras: además de las estructuras sin actividad anticonvulsiva del set A, se han incluido 13 anticonvulsivos: 11 anticonvulsivos estructuralmente similares además de carbamacepina y valpramida. THIQ-10c ($ED_{50} = 5.17 \mu\text{mol/kg}$) fue elegida como el compuesto de referencia del conjunto B. Se desea evaluar si el algoritmo sirve para diferenciar los compuestos anticonvulsivantes de los no-anticonvulsivantes, tanto en el caso de que los anticonvulsivantes sea estructuralmente diversos (conjunto A) como en el caso en que sean estructuralmente similares (conjunto B).

Se derivaron los conjuntos de n $(n-1)/2$ APs para todos los compuestos de los conjuntos A y B. Los conjuntos de pares de átomos de los compuestos de referencia, carbamacepina y THIQ-10c se presentan en la tabla 9.4.

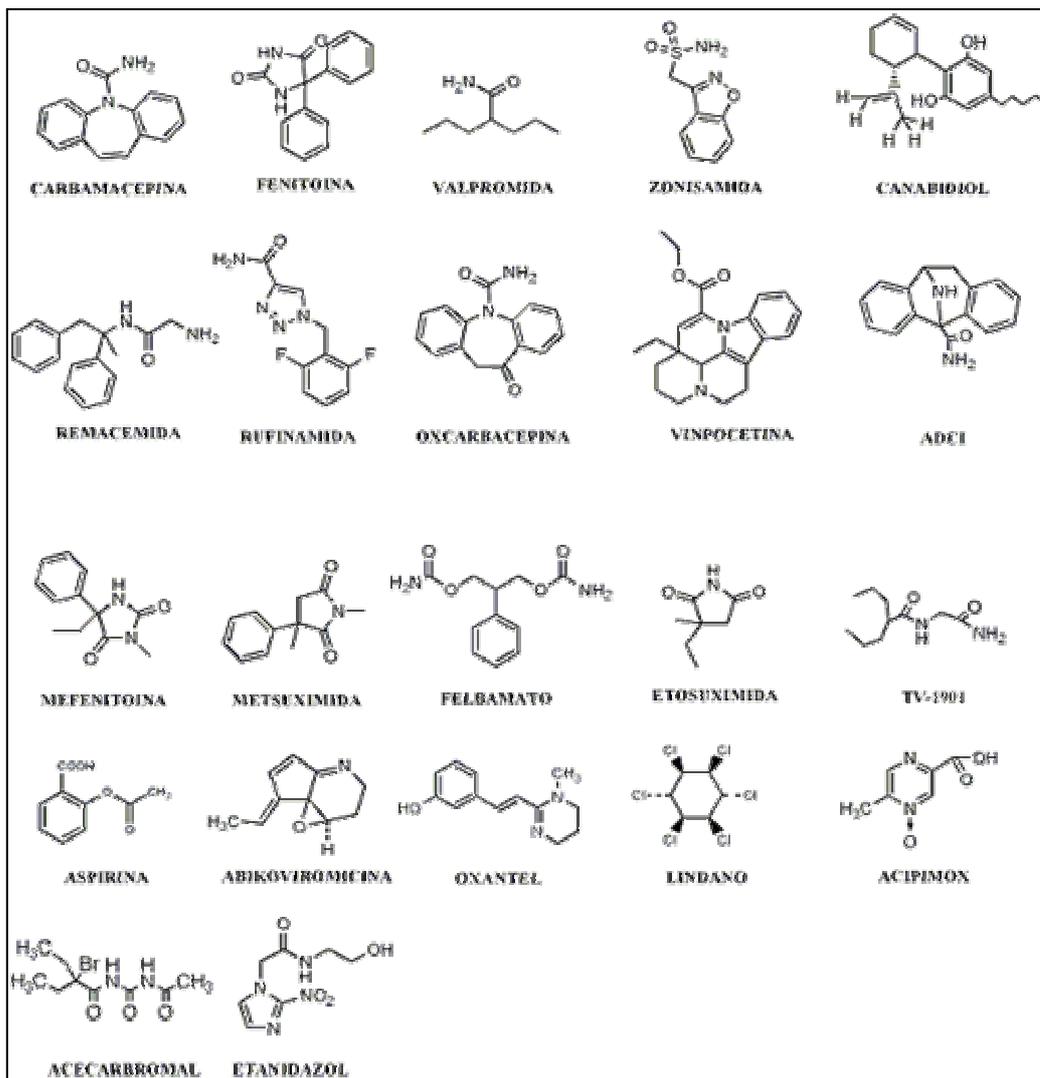


Fig. 10. Estructuras de los compuestos que componen el conjunto A.

Se calculó el valor de A en el modelo propuesto. En la tabla 9.5 se muestra un ejemplo de cálculo del factor entre corchetes del modelo cuando se compara fenitoína con carbamacepina.

Se probaron diferentes valores de b y c . A b se le asignaron valores de 0.5, 0.75 y 1; a c se le asignaron valores de 50 y 75. Una vez calculados los valores de A para las diferentes formas del modelo, se llevaron a cabo análisis de varianza para determinar si existen diferencias estadísticamente significativas entre el valor medio de A



para los compuestos activos e inactivos de cada conjunto.

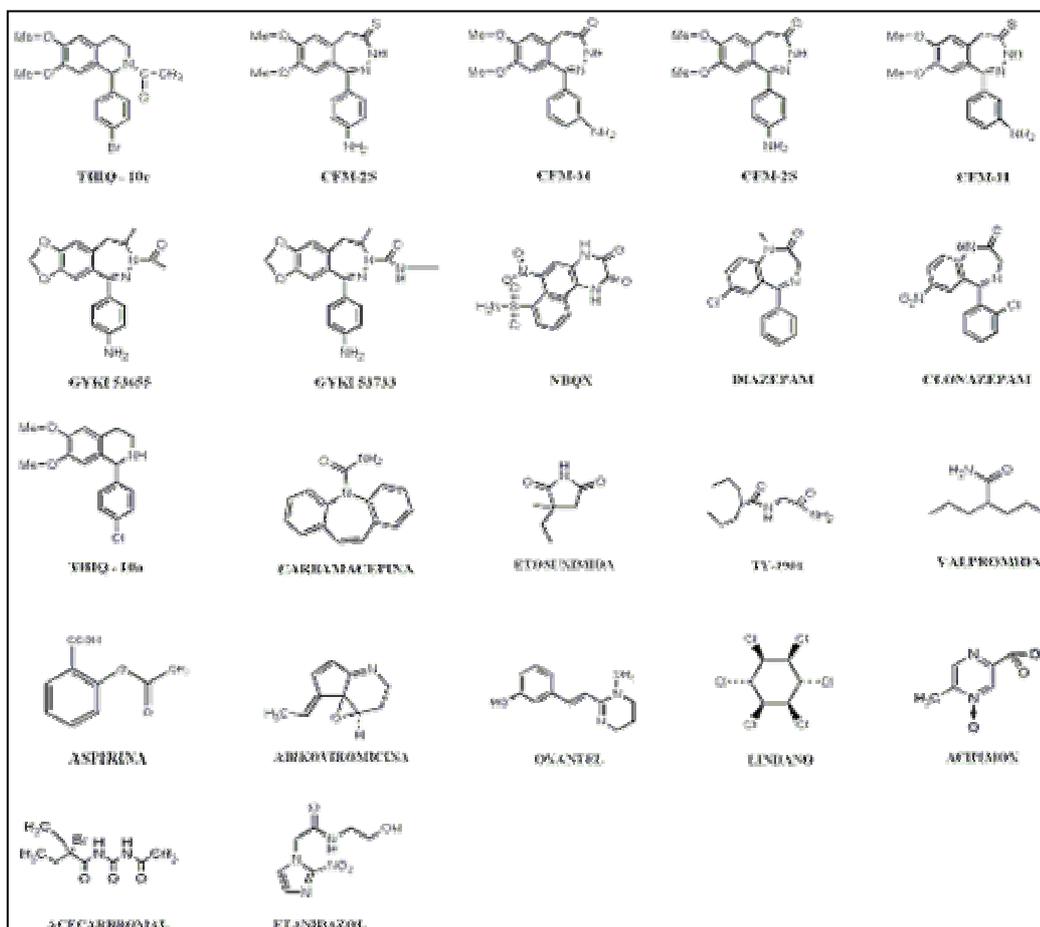


Fig. 9.11. Estructuras que componen el conjunto B.

Las tablas 9.6 y 9.7 muestran respectivamente los valores del coeficiente de Tanimoto y los valores de A obtenidos para los conjuntos A y B con una de las formas del modelo que presentó un buen desempeño en la identificación de los anticonvulsivos ($b=75$, $c=0.75$), tanto con la forma binaria como con la forma algebraica del coeficiente de Tanimoto. La tabla 9.8 muestra los resultados del test de ANOVA para las diferentes formas del modelo propuesto; también se muestran los resultados del test de ANOVA para el coeficiente de Tanimoto en sus formas binaria y algebraica. Se presentan los valores



Tabla 9.4. Conjuntos de APs derivados de las estructuras de carbamazepina y THIQ-10c.

Carbamazepine		THIQ-10c			
Count	Atom pair	Count	Atom pair	Count	Atom pair
7	C*X2-2-C*X2	2	C*X2-2-C*X2	2	CX1-5-C*X2
2	C*X3-2-C*X3	8	C*X2-2-C*X3	4	CX1-5-C*X3
6	C*X2-2-C*X3	2	C*X2-3-BrX1	1	CX1-5-CX2
1	C*X3-2-NX1	2	C*X2-3-C*X2	2	CX1-5-OX2
3	C*X3-2-NX3	8	C*X2-3-C*X3	3	CX1-6-C*X2
1	C*X3-2-O*X1	2	C*X2-3-OX2	3	CX1-6-C*X3
6	C*X2-3-C*X2	2	C*X2-4-BrX1	1	CX1-6-CX1
11	C*X2-3-C*X3	3	C*X2-4-C*X2	1	CX1-6-CX2
2	C*X2-3-NX3	3	C*X2-4-C*X3	1	CX1-6-CX3
8	C*X2-4-C*X2	3	C*X2-4-NX3	3	CX1-7-C*X2
10	C*X2-4-C*X3	2	C*X2-4-OX2	1	CX1-7-C*X3
6	C*X2-4-NX3	2	C*X2-5-C*X2	2	CX1-7-CX2
7	C*X2-5-C*X2	8	C*X2-5-C*X3	1	CX1-7-CX3
14	C*X2-5-C*X3	3	C*X2-5-NX3	1	CX1-7-NX3
2	C*X2-5-NX1	4	C*X2-6-C*X2	2	CX1-8-C*X2
2	C*X2-5-NX3	7	C*X2-6-C*X3	5	CX1-8-C*X3
2	C*X2-5-O*X1	3	C*X2-6-O*X1	1	CX1-8-CX2
7	C*X2-6-C*X2	2	C*X2-7-C*X2	1	CX1-8-NX3
8	C*X2-6-C*X3	4	C*X2-7-C*X3	1	CX1-8-OX2
6	C*X2-6-NX3	3	C*X2-7-O*X1	1	CX1-9-BrX1
6	C*X2-6-O*X1	2	C*X2-7-OX2	4	CX1-9-C*X2
7	C*X2-7-C*X2	1	C*X2-8-BrX1	1	CX1-9-C*X3
2	C*X2-7-NX1	3	C*X2-8-C*X3	1	CX1-9-CX1
2	C*X2-7-O*X1	4	C*X2-8-OX2	1	CX1-9-O*X1
3	C*X2-8-C*X2	1	C*X2-9-BrX1	1	CX1-9-OX2
4	C*X3-3-C*X3	2	C*X2-9-OX2	1	CX2-2-C*X3
2	C*X3-3-NX3	1	C*X3-10-BrX1	1	CX2-2-CX2
5	C*X3-4-C*X3	1	C*X3-10-OX2	1	CX2-2-NX3
2	C*X3-4-NX1	1	C*X3-2-BrX1	1	CX2-3-C*X2
2	C*X3-4-OX1	2	C*X3-2-C*X3	3	CX2-3-C*X3
2	C*X3-5-NX1	1	C*X3-2-NX3	1	CX2-3-CX3
2	C*X3-5-O*X1	1	C*X3-2-O*X1	1	CX2-3-NX3
1	NX1-3-O*X1	2	C*X3-2-OX2	2	CX2-4-C*X2
1	NX3-3-NX1	3	C*X3-3-C*X3	4	CX2-4-C*X3
1	NX3-3-O*X1	2	C*X3-3-NX3	1	CX2-4-CX3
7	C*X2-2-C*X2	2	C*X3-3-OX2	1	CX2-4-O*X1
2	C*X3-2-C*X3	6	C*X3-4-C*X3	3	CX2-5-C*X2
6	C*X2-2-C*X3	1	C*X3-4-NX3	3	CX2-5-C*X3
1	C*X3-2-NX1	2	C*X3-4-OX2	1	CX2-5-O*X1
3	C*X3-2-NX3	1	C*X3-5-BrX1	1	CX2-5-OX2
1	C*X3-2-O*X1	2	C*X3-5-C*X3	4	CX2-6-C*X2
6	C*X2-3-C*X2	1	C*X3-5-NX3	1	CX2-6-C*X3
11	C*X2-3-C*X3	2	C*X3-5-O*X1	2	CX2-6-OX2
2	C*X2-3-NX3	2	C*X3-5-OX2	2	CX2-7-C*X2
8	C*X2-4-C*X2	3	C*X3-6-C*X3	1	CX2-7-C*X3
10	C*X2-4-C*X3	2	C*X3-6-NX3	1	CX2-7-OX2
6	C*X2-4-NX3	1	C*X3-6-O*X1	1	CX2-8-BrX1
7	C*X2-5-C*X2	1	C*X3-6-OX2	1	CX2-9-BrX1
14	C*X2-5-C*X3	1	C*X3-7-BrX1	2	CX3-2-C*X3
2	C*X2-5-NX1	1	C*X3-7-C*X2	1	CX3-2-NX3
2	C*X2-5-NX3	2	C*X3-7-C*X3	3	CX3-3-C*X2
2	C*X2-5-O*X1	1	C*X3-7-O*X1	2	CX3-3-C*X3
7	C*X2-6-C*X2	2	C*X3-7-OX2	3	CX3-4-C*X2
8	C*X2-6-C*X3	2	C*X3-8-BrX1	1	CX3-4-C*X3
6	C*X2-6-NX3	1	C*X3-8-C*X3	1	CX3-4-O*X1
6	C*X2-6-O*X1	2	C*X3-8-O*X1	2	CX3-5-C*X3
7	C*X2-7-C*X2	1	C*X3-8-OX2	1	CX3-5-OX2
2	C*X2-7-NX1	1	C*X3-9-BrX1	1	CX3-6-BrX1
2	C*X2-7-O*X1	1	C*X3-9-C*X3	1	CX3-6-OX2
3	C*X2-8-C*X2	1	C*X3-9-OX2	1	CX3-7-C*X3
4	C*X3-3-C*X3	2	CX1-10-C*X2	1	CX3-8-C*X3
2	C*X3-3-NX3	1	CX1-10-C*X3	1	NX3-3-O*X1
5	C*X3-4-C*X3	1	CX1-10-CX1	1	NX3-6-OX2



2	C*X3-4-NX1	1	CX1-10-O*X1	1	NX3-7-BrX1
2	C*X3-4-OX1	1	CX1-11-BrX1	1	NX3-7-OX2
2	C*X3-5-NX1	1	CX1-11-C*X3	1	O*X1-9-BrX1
2	C*X3-5-O*X1	1	CX1-12-BrX1	1	OX2-10-BrX1
1	NX1-3-O*X1	1	CX1-2-C*X3	1	OX2-11-BrX1
1	NX3-3-NX1	2	CX1-2-OX2	1	OX2-4-OX2
1	NX3-3-O*X1	2	CX1-3-C*X3	1	OX2-8-O*X1
7	C*X2-2-C*X2	1	CX1-3-NX3	1	OX2-9-O*X1
2	C*X3-2-C*X3	1	CX1-3-O*X1	1	CX1-4-CX2
6	C*X2-2-C*X3	2	CX1-4-C*X2	1	CX1-4-CX3
1	C*X3-2-NX1	2	CX1-4-C*X3		

Tabla 9.5. Tipos de APs comunes a carbamacepina y fenitoína, y valor del término negativo entre corchetes del modelo para cada tipo de AP.

Tipo de par de átomos	Cantidad en carbamacepina	Cantidad en fenitoína	$\left(\frac{ni \cdot \Delta n_i}{1 + ABS(\lambda_i - 6)} \right)$ c
C*X3-2-O*X1	1	2	-0.267
C*X2-5-O*X1	2	4	-2.667
C*X2-6-O*X1	6	7	-8.000
C*X2-7-O*X1	2	6	-5.333
C*X3-4-OX1	2	4	-1.773
C*X3-5-O*X1	2	2	0.000

del estadístico F (relación entre la varianza entre grupos e intra grupos) y el nivel de significancia p. Se observa que el desempeño de varias formas del modelo es similar al desempeño del coeficiente de Tanimoto “desnudo” para el conjunto A, mientras que en el caso del conjunto B la diferencia entre los valores medios de *A* para cada categoría de compuestos es más significativa que las diferencias entre los valores medios del coeficiente de Tanimoto. Los resultados parecen indicar que el modelo propuesto es efectivo para la identificación de anticonvulsivos estructuralmente cercanos a la estructura de referencia (ya que justamente en el conjunto B la mayoría de los anticonvulsivos son estructuralmente similares). En la figura 9.12a y b se presentan los resultados del test de ANOVA para el conjunto B, utilizando el modelo con $b = 0.75$ y $c = 75$: la diferencia es más significativa cuando se usa la forma algebraica del coeficiente de Tanimoto.



Tabla 9.6. Valores del coeficiente de Tanimoto para los compuestos de los conjuntos A y B. Los compuestos inactivos se presentan en gris.

	Compuesto	Coficiente de Tanimoto – forma binaria	Coficiente de Tanimoto – forma algebraica
S E T A	ADCI	0.34	0.72
	Canabidiol	0.08	0.12
	Felbamato	0.15	0.14
	Mefenitoína	0.24	0.29
	Metsuximida	0.28	0.24
	Oxcarbacepina	0.59	0.80
	Fenitoína	0.35	0.64
	Remacemida	0.14	0.22
	Rufinamida	0.11	0.15
	Valpromida	0.05	0.002
	Zonisamida	0.21	0.20
	Abikomicina	0.09	0.07
	ABT-594	0.07	0.08
	Acecarbromal	0.03	0.02
	Acipimox	0.09	0.07
	Aspirina	0.22	0.23
	Bergapteno	0.22	0.36
	Etanidazole	0.11	0.07
	Etosuximida	0.04	0.01
	Lindano	0.00	0.00
	Metisazona	0.24	0.29
	Metocarbamol	0.07	0.11
Oxantel	0.19	0.27	
TV-1901	0.06	0.01	
S E T B	Carbamacepina	0.14	0.40
	CFM11	0.35	0.58
	CFM11S	0.13	0.55
	CFM2	0.34	0.58
	CFM2S	0.29	0.57
	Clonazepam	0.14	0.45
	Diazepam	0.24	0.52
	GYKI 57655	0.42	0.64
	GYKI 57773	0.28	0.60
	NBQX	0.10	0.28
	THIQ10a	0.45	0.69
	Valpromida	0.06	0.04
	Abikomicina	0.13	0.15
	ABT-594	0.10	0.11
	Acecarbromal	0.10	0.07
	Acipimox	0.07	0.07
	Aspirina	0.08	0.18
	Bergapteno	0.22	0.35
	Etanidazole	0.11	0.09
	Etosuximida	0.05	0.02
Lindano	0.00	0.00	
Metisazona	0.12	0.22	
Metocarbamol	0.23	0.15	
Oxantel	0.19	0.27	



Tabla 9.7. Valores de A para los compuestos de los conjuntos A y B.

	Compuesto	A (forma binaria de S)	A (forma algebraica de S)
S E T A	ADCI	0.34	0.60
	Canabidiol	0.15	0.22
	Felbamato	0.25	0.24
	Mefenitoína	0.18	0.21
	Metsuximida	0.20	0.18
	Oxcarbacepina	0.75	0.94
	Fenitoína	0.54	0.84
	Remacemida	0.16	0.23
	Rufinamida	0.19	0.24
	Valpromida	0.11	0.01
	Zonisamida	0.21	0.21
	Abikomicina	0.16	0.14
	ABT-594	0.13	0.14
	Acecarbromal	0.08	0.04
	Acipimox	0.16	0.13
	Aspirina	0.25	0.25
	Bergapteno	0.31	0.45
	Etanidazole	0.11	0.08
	Etosuximida	0.09	0.03
	Lindano	0.00	0.00
	Metisazona	0.18	0.21
	Metocarbamol	0.14	0.19
	Oxantel	0.15	0.19
TV-1901	0.11	0.03	
S E T B	Carbamacepina	0.25	0.55
	CFM11	0.42	0.61
	CFM11S	0.22	0.66
	CFM2	0.42	0.63
	CFM2S	0.40	0.67
	Clonazepam	0.22	0.54
	Diazepam	0.33	0.58
	GYKI 57655	0.55	0.75
	GYKI 57773	0.41	0.72
	NBQX	0.26	0.55
	THIQ10a	0.54	0.75
	Valpromida	0.12	0.09
	Abikomicina	0.22	0.24
	ABT-594	0.18	0.19
	Acecarbromal	0.18	0.14
	Acipimox	0.14	0.13
	Aspirina	0.15	0.26
	Bergapteno	0.34	0.49
	Etanidazole	0.18	0.16
	Etosuximida	0.10	0.06
Lindano	0.00	0.00	
Metisazona	0.19	0.29	
Metocarbamol	0.32	0.23	
Oxantel	0.26	0.34	
TV-1901	0.20	0.15	



Tabla 9.8. Valores de F y p par las diferentes formas evaluadas del modelo propuesto y para las distintas formas del coeficiente de Tanimoto.

	<i>Forma binaria de S</i>	S = Forma binaria del coeficiente de Tanimoto					
		c = 50			c = 75		
		b = 0.50	b = 0.75	b = 1.00	b = 0.50	b = 0.75	b = 1.00
SET A	F = 5.93 p = .023	F = 3.81 p = .063	F = 4.23 p = .051	F = 4.11 p = .055	F = 5.44 p = .030	F = 5.39 p = .030	F = 4.99 p = .036
SET B	F = 9.90 p = .005	F = 13.58 p = .001	F = 13.22 p = .001	F = 12.25 p = .002	F = 12.54 p = .002	F = 12.12 p = .002	F = 11.49 p = .0025
	<i>Forma algebraica de S)</i>	S = Forma algebraica del coeficiente de Tanimoto					
		c = 50			c = 75		
		b = 0.50	b = 0.75	b = 1.00	b = 0.50	b = 0.75	b = 1.00
SET A	F = 5.67 p = .026	F = 4.13 p = .054	F = 4.51 p = .045	F = 4.44 p = .047	F = 5.16 p = .033	F = 5.30 p = .031	F = 5.06 p = .035
SET B	F = 38.19 p < .0000	F = 34.66 p < .0000	F = 42.35 p < .0000	F = 46.33 p < .0000	F = 34.24 p < .0000	F = 41.04 p < .0000	F = 44.59 p < .0000

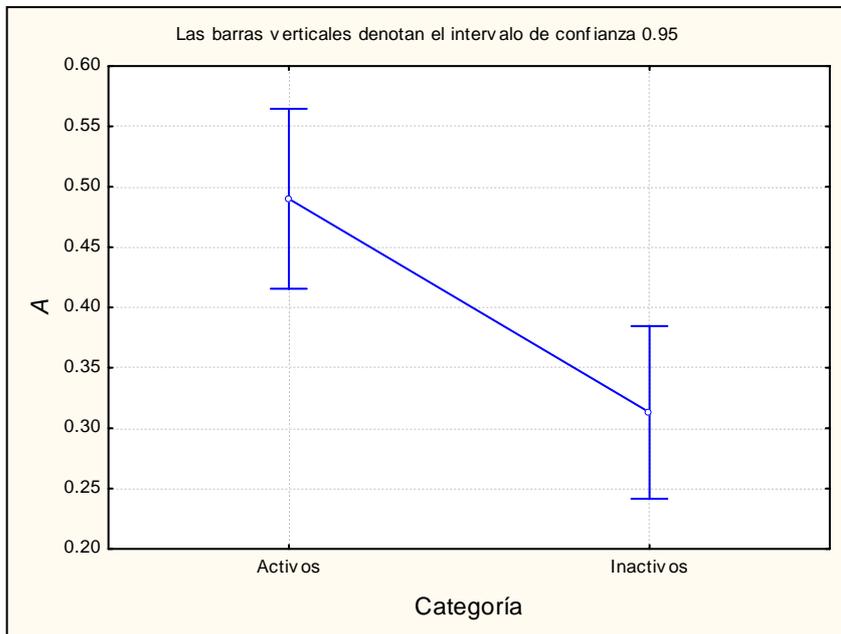


Fig. 9.12a. Resultados del ensayo de ANOVA para el conjunto B ($b=0.75$ y $c=75$), con el coeficiente de Tanimoto en su forma binaria.

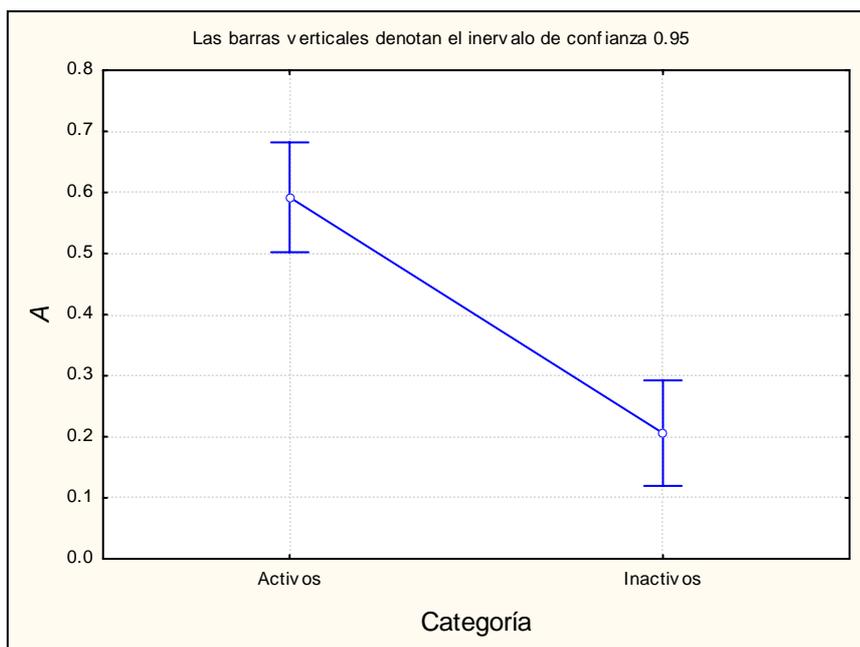


Fig. 9.12a. Resultados del ensayo de ANOVA para el conjunto B ($b=0.75$ y $c=75$), con el coeficiente de Tanimoto en su forma algebraica.

El modelo es, por lo tanto, adecuado para la búsqueda de nuevas estructuras activas en bases de datos mediante SV basado en similitud.

Referencias

1. Gavernet, L.; Domínguez-Cabrera, M.J.; Bruno-Blanch, L.E.; Estiú, G.L. 3D QSAR design of novel antiepileptic sulfamides. *Bioorg. Med. Chem.* **2007**, *15*, págs. 1556-1567
2. Design, synthesis and anticonvulsant activity of some sulfamides. *Bioorg. Med. Chem.* **2007**, *15*, págs. 5604-5614
3. Tasso, S.M.; Moon, S.C.; Bruno-Blanch, L.E.; Estiú, G.L. Characterization of anticonvulsant profile of valpromide derivatives. **2004**, *12*, págs. 3857-3869
4. Sowada, V.J. Darstellung 1,3-disubstituierter Schwefelsäure-diamide aus Sulfurylchlorid und primären Aminen. *J. Prakt. Chem.* **1963**, *4*, págs. 310-319
5. DuBois, G. Amination of aryl sulfamate esters. A convenient general synthesis of aliphatic sulfamides. *J. Org. Chem.* **1980**, *45*, págs. 5373-5375
6. O'Connell, J.F.; Rapoport, H. 1-(Benzenesulfonyl)- and 1-(p-Toluensulfonyl)-3-methylimidazolium triflates: efficient reagents



- for the preparation of arylsulfonamides and arylsulfonates. *J. Org. Chem.* **1992**, 57, págs. 4775-4777
7. Saha, A.K.; Schultz, P.; Rapoport, H. 1,1'-Carbonylbis(3-methylimidazolium) triflate: an efficient reagent for aminoacylations. *J. Am. Chem. Soc.* **1989**, 111, págs. 4856-4859
 8. Winum, J.; Toupet, L.; Barragan, V.; Dewynter, G.; Montero, J. N-(tert-Butoxycarbonyl)-N-[4-(dimethylazaniumylidene)-1,4-dihydropyridin-1-ylsulfonyl]azanide: A New Sulfamoylating Agent. Structure and Reactivity toward Amines. *Org. Lett.* **2001**, 3, 2241-2243
 9. Lee, H.K.; Bang, M.; Pak, C. S. Efficient synthesis of arylsulfamides by reaction of amines with arylsulfamoyl imidazolium triflate. **2005**, 46, págs. 7139-7142
 10. Veber, D.F.; Johnson, S.R.; Cheng, H.Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, 45, págs. 2615-2623
 11. InterBioScreen's NC Collection. <http://www.interbioscreen.com/>
 12. Newman, D.J.; Cragg, G.M.; Snader, K.M. Natural products as sources of new drugs in the period 1981-2002. *J. Nat. Prod.* **2003**, 66, págs. 1022-1037
 13. Kohen, F.E.; Carter, G.T. The evolving role of natural products in drug discovery. *Nat. Rev. Drug. Discov.* **2005**, 4, págs. 206-220
 14. Liu, H.; Li, Y.; Song, M.; Tan, X.; Cheng, D.; Zheng, S.; Shen, J.; Luo, X.; Ji, R.; Yue, J.; Hu, G.; Jiang, H.; Chen, K. Structure-based discovery of potassium channel blockers from natural products: virtual screening and electrophysiological assay testing. *Chem. Biol.* **2003**, 10, págs. 1103-1113
 15. Shen, J.; Xu, X.; Cheng, F.; Liu, H.; Luo, X.; Shen, J.; Chen, K.; Zhao, W.; Shen, X.; Jiang, H. Virtual Screening on Natural Products for Discovering Active Compounds and Target Information. *Curr. Med Chem.* **2003**, 10, págs. 2327-2342
 16. Tripos Inc., SYBYL 6.6, Saint Louis, Estados Unidos, **1999**.
 17. Carhart, R.E.; Smith, D.H.; Venkataraghavan, R. Atom pairs as molecular features in structure-activity studies: definitions and applications. *J. Chem. Inf. Comput. Sci.* **1985**, 25, págs. 64-73



Sumario y conclusiones.

*“La felicidad consiste en poder unir
el fin con el principio”*

Pitágoras

El objetivo general del presente trabajo de tesis fue el desarrollo de la Topología Molecular en el ámbito de la Facultad de Ciencias Exactas. Específicamente, esta metodología se aplicó, luego de su comprensión, en la búsqueda racional de nuevos prototipos activos en el ensayo MES en bases de datos de estructuras químicas. A fin de realizar este objetivo general el plan de trabajo incluía varios objetivos particulares: la generación de una base de datos del orden de 10^5 estructuras químicas, la generación de modelos basados en descriptores topológicos para la identificación de anticonvulsivantes novedosos mediante aplicación de screening virtual en la base de datos generada, y la aplicación de filtros ADMET para obtener estructuras no sólo activas sino también biodisponibles y seguras. Podemos afirmar que se han logrado los objetivos general y particulares propuestos:

- Se identificaron y ensayaron cinco agentes anticonvulsivos novedosos, estructuralmente diversos. El propil y el metilparabeno se han usado y usan extensamente como conservantes en alimentos, medicamentos y cosméticos, siendo, según lo indica un gran número de estudios en literatura, seguros para su administración en humanos a dosis mucho mayores a las identificadas como anticonvulsivas en el presente trabajo. El acesulfame ha sido utilizado como edulcorante en bebidas dietéticas y otros alimentos desde hace casi dos décadas. Estos datos indican que, como consecuencia de la probada seguridad de los mismos. es posible desarrollar, a partir de los



anticonvulsivantes identificados, nuevos medicamentos antiepilépticos efectivos y seguros en un tiempo relativamente breve. Ésta constituye una de las grandes ventajas del screening virtual: la identificación de nuevas actividades en fármacos ya conocidos, permitiendo responder con relativa rapidez a necesidades terapéuticas urgentes. Las estructuras ensayadas, por su novedad estructural respecto a los fármacos antiepilépticos conocidos, pueden considerarse cabezas de serie sobre las que realizar, en un futuro, modificaciones estructurales orientadas a mejorar el perfil farmacológico y farmacocinético.

- Se ha generado una base de datos de casi 470,000 compuestos y los valores correspondientes de 877 descriptores de baja dimensionalidad. Esta base de datos está disponible en el laboratorio de Química Medicinal de la Facultad de Ciencias Exactas, Universidad Nacional de La Plata, para desarrollar en un futuro nuevas aplicaciones de la metodología utilizada en la presente tesis, abordando objetivos terapéuticos distintos. En ese sentido la base ya ha sido utilizada en estudios publicados para la identificación de nuevos fármacos antichagásicos y antiepileptogénicos.
- Se ha generado y validado una función discriminante basada en descriptores topológicos, capaz de identificar compuestos antiepilepticos. Se han generado asimismo filtros particulares ADMET para predicción de a) la biodisponibilidad a nivel del sistema nervioso central y b) la capacidad para producir ataxia. La aplicación de la función discriminante y los filtros particulares conjuntamente con filtros ADMET generales ha



permitido identificar 1,740 estructuras químicas potencialmente antiepilépticas, biodisponibles y seguras de la base de datos.

- Se han detectado, en una aplicación original del screening virtual, siete nuevos transferidores del grupo sulfurilo. Uno de ellos, el acesulfame, se ha utilizado para sintetizar una sulfamida anticonvulsiva mediante una reacción novedosa. Las características del acesulfame le brindan varias ventajas respecto a otros transferidores ya conocidos (menor reactividad, mayor seguridad ambiental). Esta parte del trabajo, en fase incipiente de desarrollo, ha sido llevado a cabo en conjunto con el Lic. Juan Elvira, como parte de su tesis doctoral orientada a la búsqueda de rutas sintéticas alternativas de sulfamidas anticonvulsivantes. Actualmente se está realizando la optimización de la reacción para favorecer la formación de la N,N'-dibencilsulfamida. Adicionalmente, otros integrantes del equipo del trabajo han iniciado la planificación de síntesis de otras sulfamidas a partir del acesulfame.
- Se ha desarrollado, conjuntamente con la Dra. Luciana Gavernet una metodología combinada QSAR 2D-3D, que integra dos líneas desarrolladas en el equipo de trabajo para optimizar la identificación de nuevos agentes anticonvulsivos. Esta metodología utiliza en forma combinada la función discriminante topológica, los filtros ADMET generales y un farmacóforo 3D para la identificación de drogas antiMES. El protocolo desarrollado está de acuerdo con las tendencias más actuales de screening virtual, que proponen el uso sinérgico de las metodologías 2D y 3D, aprovechando conjuntamente las ventajas particulares de cada una.



Debe destacarse que se ha trabajado en forma interdisciplinaria con otros miembros del grupo de trabajo, abordando integralmente los diversos aspectos implicados en el desarrollo de nuevos fármacos: búsqueda computacional, síntesis y ensayos farmacológicos. Pese a que el trabajo realizado en esta tesis es fundamentalmente teórico, los resultados han sido validados mediante ensayos farmacológicos (en el caso de los anticonvulsivantes evaluados) y síntesis (en el caso del acesulfame como transferidor del grupo sulfurilo).

El presente trabajo de tesis demuestra la potencialidad de técnicas sencillas y computacionalmente poco costosas, tales como el análisis lineal discriminante y el screening virtual basado en topología molecular, en la identificación de nuevas estructuras líderes. Se ha verificado que el screening virtual 2D constituye una herramienta rápida, efectiva, que requiere poca inversión en equipamiento para el desarrollo de nuevos agentes terapéuticos, que además se halla de acuerdo con los principios bioéticos que exigen optimizar el uso de los ensayos preclínicos y clínicos siempre que sea posible.

Durante este trabajo de tesis se han concretado 12 presentaciones a congresos (5 nacionales y 7 regionales e internacionales) y 8 publicaciones en revistas científicas con referato (6 internacionales y 2 nacionales), además de 2 trabajos aceptados para su publicación en revistas internacionales y 2 trabajos en redacción. Se ha obtenido también, mediante concurso, un subsidio de viaje de la Federación Europea de Química Medicinal (EFMC) para asistir al XIX Simposio Internacional de Química Medicinal (Turquía, Agosto-Septiembre de 2006) a presentar un trabajo vinculado al núcleo de esta tesis. El trabajo referido al descubrimiento de la actividad anticonvulsiva del ácido abiético ha sido seleccionado para su difusión a través del servicio LeadDiscovery's DailyUpdates por considerarse de interés en



el área de desarrollo de nuevos fármacos. Las publicaciones se detallan a continuación:

- Talevi, A.; Sella-Craverio, M.; Castro, E.A.; Bruno-Blanch, L.E. Discovery of anticonvulsant activity of abietic acid through application of linear discriminant analysis. *Bioorg. Med. Chem. Lett.* **2007**, 17, 1684-1690 (letter)
- Talevi, A.; Prieto, J.J.; Bruno-Blanch, L.E.; Castro, E.A. New similarity-based algorithm and its application to classification of anticonvulsant compounds. *J. Enz. Inhib. Med. Chem.* **2007**, 22, págs. 253-265 (full paper)
- Duchowicz, P.R.; Talevi, A.; Bellera, C.; Bruno-Blanch, L.E.; Castro, E.A. Application of descriptors based on Lipinski's rules in the QSPR study of aqueous solubilities. *Bioorg. Med. Chem.* **2007**, 15, págs. 3711-3719 (full paper)
- Bellera, C.L.; Talevi, A.; Bruno-Blanch, L.E. Aplicación de análisis lineal discriminante en la búsqueda de drogas antiepiléptogénicas. *Lat. Am. J. Pharm.* **2007**, 26, págs. 244-252 (full paper)
- Talevi, A.; Bellera, C.L.; Castro, E.A. Bruno-Blanch, L.E. Application of molecular topology in descriptor-based virtual screening for the discovery of new anticonvulsant agents. *Drug Fut.* **2007**, 31 (Suppl. A), pág. 188 (resumen)
- Talevi, A.; Castro, E.A.; Bruno-Blanch, L.E. New solubility models based on descriptors derived from the detour matrix. *J. Arg. Chem. Soc.* **2006**, 94, págs. 129-141 (full paper)
- Talevi, A.; Prieto, J.J.; Bruno-Blanch, L.E.; Castro, E.A. On application of Carhart atom pairs to predict anticonvulsant activity. *Am. J. Biochem. Biotechnol.* **2006**, 2, págs. 199-128 (full paper)
- Prieto, J.J.; Talevi, A.; Bruno-Blanch, L.E. Application of linear discriminant analysis in the virtual screening of antichagasic drugs through trypanothione reductase inhibition. *Mol. Divers.* **2006**, 10, págs. 361-375 (full paper)



Trabajos aceptados.

- Talevi, A.; Bellera, C.L.; Castro, E.A.; Bruno-Blanch, L.E. A Successful Virtual Screening Application: Prediction of Anticonvulsant Activity in MES test of Widely Used Pharmaceutical and Food Preservatives Methylparaben and Propylparaben. *J. Comput. Aid. Mol. Des.* (full paper)
- Gavernet, L.; Talevi, A.; Castro, E.A.; Bruno-Blanch, L.E. Combined Virtual Screening 2D and 3D QSAR Methodology for the Selection of New Anticonvulsant Candidates from a Natural Product Library. *QSAR and Comb. Sci.* (full paper)

Trabajos en redacción.

- Elvira, J.E.; Talevi, A.; Bellera, C.L.; Bruno-Blanch, L.E. Discovery of New Sulfuryl Transfer Agent: Acesulfame - An original application of virtual screening (full paper)
- Talevi, A.; Sella-Craverro, M.; Castro, E.A.; Bruno-Blanch, L.E. Anticonvulsant profile of new anticonvulsant agent propylparaben (letter)

“¿Y qué hay con eso? ¿Qué pasa con estar en presencia de la nada? Hubo otros hombres que también estuvieron en presencia de la nada y tuvieron el coraje de encerrarla en un círculo y denominarlo ‘cero’.

Fue ese cero el que permitió el vuelo de las matemáticas y con ello uno de los más importantes vuelos del pensamiento humano.

Los hombres que crearon el número cero no tenían nada, pero tenían los silencios, los lugares vacíos, el aire y todo aquello casi tan inexistente como el olvido. Tenían la facultad de atrapar, de aprehender con las manos abiertas innumerables cosas y entre esas cosas la ausencia de todas las cosas encerrada en un círculo de tinta”

Dalmiro Sáenz, “Carta corregida a mi futura ex-mujer”