

Técnicas QSAR en Diseño de Fármacos

M. PASTOR Y J. ALVAREZ-BUILLA

Departamento de Química Orgánica. Universidad de Alcalá.

E-28871 Alcalá de Henares. Madrid.

1.0 INTRODUCCION.

Los denominados métodos QSAR (relaciones cuantitativas estructura-actividad) también llamados métodos LFER (relaciones lineales de energía libre), han agrupado históricamente a todas las técnicas que han intentado establecer modelos empíricos de comportamiento sobre familias de compuestos biológicamente activos, en orden a obtener, de la manera mas eficaz posible, óptimos de actividad, a partir de los datos de comportamiento de un número limitado de productos.

La evolución de este tipo de métodos ha sido recogida por Craig en una revisión reciente (Craig, 1984). La primera vez que se habló de un modelo "aditivo" de relaciones estructura- actividad, tal como hoy lo entendemos, fue en un artículo de Bruice, Karasch y Winzler (Bruice, 1956), donde los autores describían la utilización de un modelo *de novo*. Por desgracia, el artículo se publicó en una revista de baja difusión, lo que hizo que no tuviera eco en los años siguientes. Pocos años después, Free y Wilson (Free, 1964) publicaron esencialmente el mismo método, sin que ni ellos ni los editores de la revista fueran capaces de detectar el trabajo anterior, y así el modelo ha sido conocido desde entonces como de Free-Wilson.

De manera simplificada, la técnica está basada en el uso de series de compuestos derivados de una estructura común, en los que se observan las alteraciones producidas en su actividad biológica en función de la presencia de sustituyentes diversos. A partir de estos datos, es posible calcular coeficientes aditivos *de novo* para cada sustituyente, con lo que es posible diseñar la mejor combinación de sustituyentes para alcanzar el óptimo de actividad.

Simultáneamente con el trabajo de Free y Wilson, otro trabajo fundamental para el desarrollo de las técnicas QSAR se estaba produciendo, por el grupo de Corwin Hansch, en el Pomona College (Claremont, California). C. Hansch era entonces un investigador con formación en química orgánica física, y estaba familiarizado con el trabajo de Hammett. El y Toshio Fujita, que fue a Pomona de año sabático a principios de los años sesenta, pusieron a punto el denominado modelo paramétrico -también

llamado modelo de Hansch- en el que se correlaciona la actividad biológica de series de productos con sus propiedades físico-químicas (Hansch, 1964). Ambos métodos constituyen la base para la optimización de actividades a partir de prototipos cabeza de serie, tal como se utiliza en la actualidad.

La metodología QSAR sigue, independientemente del modelo que se utilice, unos pasos comunes que se hallan sintetizados en la Fig. 1.1. El punto de partida en todos los casos implica la existencia de un **prototipo cabeza de serie**, que se define como un producto que muestra actividad en relación con

METODOLOGIA QSAR

PUNTO DE PARTIDA

PROTOTIPO CABEZA DE SERIE

FASE DE APRENDIZAJE

SERIE DE EXPLORACION

IDENTIFICACION
SUSTITUYENTES

ACTIVIDAD
BIOLÓGICA

X_i

A

ANALISIS DE REGRESION

FASE DE OPTIMIZACION

MODELO $A=f(X_i)+cte$

PRODUCTOS OPTIMOS

Fig. 1.1

el objetivo terapéutico buscado. Este producto, aunque reviste interés como modelo a desarrollar, no puede habitualmente ser utilizado como fármaco - debido a que tiene una actividad escasa, o una toxicidad elevada, o características farmacocinéticas inadecuadas.- y ha de ser optimizado en sus propiedades para pasar a ser un candidato a fármaco.

Cuando se dispone de un prototipo, es preciso diseñar una serie de exploración, que está constituida por un conjunto de productos análogos del prototipo, que permitan el establecimiento de las primeras relaciones estructura-actividad. Los miembros de la serie de exploración han de ser sintetizados y analizados en su actividad biológica.

Los miembros de la serie de exploración están constituidos por un núcleo común, y unos sustituyentes o fragmentos variables, que son característicos de cada producto de la serie. Dichos sustituyentes variables han de ser identificados por descriptores, que serán utilizados como variables independientes (X_i) en el modelo. Los descriptores son binarios -indicando simplemente presencia o ausencia del fragmento- en los modelos *de novo*, y son mayoritariamente escalares -es decir con valores numéricos diversos- en los modelos paramétricos. Por otra parte, los valores de actividad biológica -normalmente expresados en forma logarítmica- se utilizan como variable dependiente (A) en el modelo.

Una vez establecidos ambos conjuntos de valores X_i y A , se utilizan técnicas de regresión múltiple para obtener un modelo de la forma indicada en la eq. 1.1, donde la actividad está expresada en forma de una constante mas las contribuciones de los sustituyentes variables utilizados.

$$A = f(X_i) + \text{cte.} \quad (\text{Eq. 1.1})$$

El modelo ha de ser analizado en su calidad estadística, para evaluar su capacidad de predicción, y frecuentemente es utilizado -en una denominada fase de comprobación- para calcular la actividad de miembros de la familia de productos analizada. Cuanta mayor calidad estadística tenga el modelo, tanto mas fiables serán las predicciones de actividad obtenidas.

La fase de optimización implica que una vez obtenido un modelo de buena calidad, es posible calcular los productos con actividad óptima dentro de la familia estudiada, que habrán sido obtenidos con un mínimo de esfuerzo experimental. Por supuesto, esto no se consigue al primer intento, de hecho, el proceso de optimización de un prototipo es un proceso cíclico (Fig. 1.2), en el que la serie de exploración inicial produce un modelo de baja calidad, que es refinado por el diseño y la síntesis de series complementarias, hasta obtener un modelo de calidad suficiente para ser utilizado en el diseño de los productos óptimos.

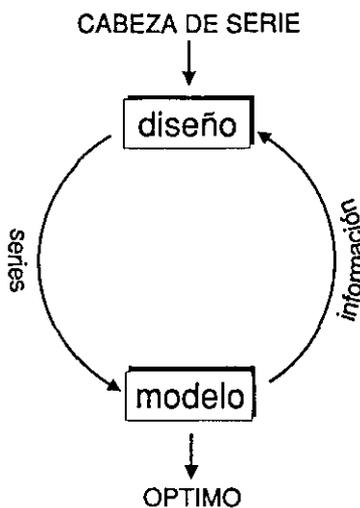


Fig. 1.2

2.0 MODELOS DE NOVO

2.1 La metodología de Free-Wilson (Kubinyi, 1990).

Todos los químicos implicados en desarrollo de fármacos siguen una metodología similar a la descrita en la Fig. 1.1. Partiendo de un prototipo, como el carbazol sustituido de la Fig. 2.1, se obtienen series de productos con

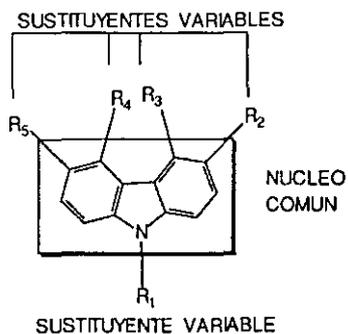


Fig. 2.1

modificaciones estructurales progresivas. Cada modificación puntual realizada produce un cambio definido en la actividad biológica, y como

primera aproximación, este cambio es fijo e independiente de las restantes modificaciones introducidas en la molécula. Expresado en términos matemáticos, la actividad biológica A , debe ser -en una familia de productos derivados del prototipo- función de la contribución del núcleo común a toda la serie (μ), y de las contribuciones (a_i) de los sustituyentes variables (X_i , variable binaria, que toma valor 1 cuando el sustituyente está presente en la molécula, y valor 0 cuando no está). El modelo de comportamiento está representado por la eq. 2.1, donde A es la actividad (Free, 1964) -representada en forma logarítmica ($\log A$ ó $\log 1/C$) en el modelo de Fujita-Ban, más reciente (Fujita, 1971)- y μ es, bien la media de la actividad de los productos de la serie (Free, 1964), bien la actividad, calculada por el modelo de la contribución del núcleo común sin sustituyentes.

$$A = \sum a_i + \mu \quad (\text{Eq. 2.1})$$

El modelo de Free-Wilson en todas sus variantes es el único modelo auténtico de relaciones estructura-actividad ya que la contribución a la actividad de cada fragmento se calcula directamente por regresión múltiple. En la aproximación extratermodinámica se correlacionan propiedades fisico-químicas con la actividad, y no directamente fragmentos estructurales.

Por otra parte, la utilidad del análisis de Free-Wilson es más limitada que la de los modelos extratermodinámicos. Mientras que éstos producen conclusiones generales acerca de la acción de un fármaco a nivel molecular, los modelos *de novo* permiten sólo una cuantificación de los efectos que los cambios estructurales producen sobre la actividad biológica.

2.2 El diseño de la serie de exploración.

Para poder utilizar técnicas QSAR en diseño de fármacos, el primer requisito es cuidar el diseño de la serie de exploración, en orden a conseguir que los productos sean congenéricos (Rekker, 1985), lo que significa: a) que todos los productos tengan el mismo esqueleto común, y b) que los sustituyentes se influencien sólo de manera aditiva -es decir que no interaccionen por puentes de hidrógeno, por conjugación o por efectos estéricos-.

Dado que en desarrollo de fármacos el coste del trabajo experimental -síntesis y análisis farmacológico- es mucho más caro que el tratamiento estadístico de los datos obtenidos, es importante conseguir un diseño óptimo de la serie de exploración, para evitar las limitaciones derivadas de la falta de calidad de la muestra (Hudson, 1970).

El número mínimo de productos requeridos para un análisis de Free-Wilson es de uno por cada parámetro de la ecuación de regresión -uno para el producto de referencia, y uno por cada sustituyente diferente del correspondiente sobre el producto de referencia. Si hay j posiciones diferentes en de variación estructural, con n_j sustituyentes diferentes para cada posición, el mínimo número de análogos viene dado por la ecuación 2.2 (Martin, 1978), mientras que el número de análogos viene dado por la ecuación 2.3.

$$N_{\min} = \sum_j (n_j - 1) + 1 \quad (\text{Eq. 2.2})$$

$$N_{\max} = n_1 \cdot n_2 \cdot n_3 \dots n_{j-1} \cdot n_j \quad (\text{Eq. 2.3})$$

Así, para una serie con tres posiciones de sustitución en el prototipo, y cinco grupos diferentes para cada posición, el mínimo número de análogos para obtener un modelo de Free-Wilson será de $4+4+4+1 = 13$ mientras que el número de derivados posibles será de $5 \times 5 \times 5 = 125$. El número de análogos seleccionados para la síntesis y análisis farmacológico será siempre el resultado de un compromiso entre la accesibilidad sintética y los requerimientos estadísticos. Cuanto mayor sea la serie de exploración, mayor será el número de grados de libertad y mas fiables serán las predicciones del modelo final, pero de manera simultánea los costes y el esfuerzo aumentan con el número de análogos incluidos en dicha serie.

Todas las recomendaciones sobre el número de grados de libertad requeridos y sobre el número de análogos por cada parámetro son arbitrarios (Kubinyi, 1990), porque dependen del diseño de la serie, de que los sustituyentes estén distribuidos uniformemente o no, de la calidad de los datos de actividad biológica, etc. Normalmente, cada fragmento debe repetirse entre dos y cuatro veces en la serie, y deben estar presentes en frecuencias comparables. Dos tipos de problemas son frecuentes y deben ser evitados en el diseño de una serie de exploración: a) si un sustituyente aparece una sola vez en la serie, se produce una determinación de punto único, sin significación estadística y b) si dos sustituyentes aparecen siempre juntos, no es posible calcular la contribución individual de cada uno (Purcell, 1973) y deben ser tratados como un único sustituyente (Kubinyi, 1976) para evitar fracasos en el análisis de regresión por dependencia lineal en la matriz de sustituyentes.

Para obtener una serie de exploración de buena calidad con el mínimo número de derivados se han propuesto varios métodos. V. Austel ha descrito uno manual basado en el diseño 2ⁿ factorial (Austel, 1985), y J.

Elguero y col. han descrito un método cuantitativo para extraer un conjunto óptimo de todos los análogos posibles, basado en la maximización del determinante de la matriz de sustituyentes (Cativiela, 1983).

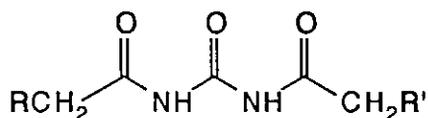
2.3 La elaboración de un modelo *de novo*.

Dejando aparte las versiones iniciales del modelo de Free-Wilson (Kubinyi, 1990), la modificación denominada de Fujita-Ban (Fujita, 1971) está basada en la utilización del modelo indicado en la ecuación 2.4, donde la constante μ está definida como la actividad calculada del producto no sustituido de la serie. Todas las contribuciones a la actividad b_{jk} están referidas al hidrógeno en la misma posición de sustitución.

$$\log A = \sum b_{jk}X_{jk} + \mu \quad (\text{Ec.2.4})$$

La metodología se muestra con facilidad sobre un ejemplo muy esquemático, desarrollado a partir de ureas substituidas con actividad hipnótica (Kubinyi, 1976a). La matriz de sustituyentes está representada en la Tabla 2.1, siendo una matriz peculiar, puesto que los valores que cada variable puede tomar serán 0, 1, ó 2, debido a la existencia de productos con sustitución simétrica. Para simplificar el cálculo de regresión, es necesario eliminar la dependencia lineal entre columnas relacionadas de la matriz -que son las referidas a sustituyentes alternativos sobre una misma posición- por lo que se elimina una entre cada conjunto de columnas relacionadas, como se indica en la columna de la variable X_1 , que corresponde al sustituyente mas sencillo (H), cuya contribución se iguala a 0. Así pues, el producto con dos sustituyentes H -el diacetil derivado- es el que se toma como patrón, y su actividad determina el valor de μ en el modelo.

Tabla 2.1. Matriz de Fujita-Ban de N,N'-diacilureas con actividad hipnótica.



Comp	H	Me	Et	Pro	But	Pent	N ₁ ,N ₂	Log (1/C)
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆		
1	1	1	0	0	0	0	6	1,84
2	0	2	0	0	0	0	9	2,06
3	1	0	1	0	0	0	8	2,16
4	0	1	1	0	0	0	12	2,23
5	1	0	0	1	0	0	10	2,27
6	0	0	0	2	0	0	16	2,40
7	0	1	0	1	0	0	15	2,35
8	1	0	0	0	1	0	12	2,46
9	0	0	1	1	0	0	20	2,38
10	0	1	0	0	1	0	18	2,25
11	1	0	0	0	0	1	14	2,55
12	0	0	0	2	0	0	25	2,32
13	0	0	1	0	1	0	24	2,28

Nota.- El término de interacción N.N' se obtiene multiplicando el número de carbonos de los dos restos acilo en cada molécula.

La regresión múltiple sobre la Tabla 2.1, produce un modelo como el indicado en la ecuación 2.5, en el que aparecen todos los parámetros estadísticos habituales en los modelos QSAR. Lo primero de todo es que los coeficientes de las variables binarias han de presentarse indicando entre paréntesis los intervalos de confianza al 95%. Asimismo, la calidad del modelo empírico se define por medio del coeficiente de correlación, r^2 - a veces se utiliza r , según los autores- la desviación standard s , y el valor F (Kubinyi, 1990). El coeficiente de correlación es una medida relativa del ajuste de los datos al modelo, calculándose a partir de la varianza residual - varianza no explicada por el modelo- y la varianza total de los datos. Cuanto menor sea la varianza no explicada, tanto mayor será r^2 , tendiendo su valor a 1. La desviación standard s es una medida absoluta del ajuste de los datos al modelo, dependiendo solamente de la varianza no explicada y del número de grados de libertad del modelo. El valor tiende a cero en los modelos de

mejor calidad. La significación de un modelo de regresión se prueba por su valor F, que correlaciona la varianza explicada (r^2) por el número de grados de libertad, con la varianza no explicada ($1-r^2$) por el número de variables del modelo. Cuanto mas alto es el porcentaje de varianza explicada por el modelo, mayor será el valor de F, mientras que la existencia de variables con baja contribución a la explicación de la varianza, tenderán a disminuir dicho valor, que tiende a infinito en los mejores modelos.

Volviendo a la ecuación 2.5, es posible comprobar como explica un pequeño porcentaje de la varianza observada, que mejora ostensiblemente cuando se incorpora a la regresión un término de interacción no lineal de Bocek-Kopecky (Bocek, 1964) (Kubinyi, 1990) $N_1 \cdot N_2$, que representa el producto del número de carbonos presentes en ambos restos acilo de cada producto, y que está diseñado arbitrariamente para representar efectos no lineales dependientes de la longitud de cadena de los grupos acilo. En el segundo modelo -ecuación 2.6- los parámetros estadísticos son claramente de mejor calidad, y el citado termino de interacción tiene una contribución negativa a la actividad, a medida que aumenta la longitud de las cadenas. La variable X_6 se marca con un asterisco, debido a que su contribución está calculada a partir de un único producto, por lo que su valor es solamente orientativo.

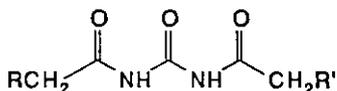
$$\begin{aligned} \text{Log } 1/\text{IC} = & 2,039 (\pm 0,34) - 0,003 (\pm 0,22) X_2 + 0,149 (\pm 0,22) X_3 + 0,186 (\pm 0,23) \\ & X_4 + 0,242 (\pm 0,29) X_5 + 0,511 (\pm 0,47) X_6^* \quad (\text{Ec. 2.5}) \\ n = 13 \quad r^2 = & 0,676 \quad s = 0,135 \quad F = 2,93 \end{aligned}$$

$$\begin{aligned} \text{Log } 1/\text{IC} = & 1,930 (\pm 0,13) + 0,430 (\pm 0,17) X_2 + 0,910 (\pm 0,28) X_3 + 1,249 \\ & (\pm 0,39) X_4 + 1,487 (\pm 0,45) X_5 + 1,813 (\pm 0,49) X_6^* - 0,085 (\pm 0,03) N_1 \cdot N_2 \\ & \quad (\text{Ec. 2.6}) \\ n = 13 \quad r^2 = & 0,964 \quad s = 0,049 \quad F = 26,59 \end{aligned}$$

Una vez establecido el modelo, es fácil conocer la contribución de cada fragmento considerado, que corresponde al valor del coeficiente de cada variable X. Así, considerando el modelo 2.6 es fácil comprobar como existe, en el intervalo estudiado, una relación lineal entre la contribución de un fragmento y su longitud, resultando el de contribución mayor el X_6 -con las limitaciones citadas en relación con este valor- que corresponde a la presencia de un grupo heptanoilo, que tiene un coeficiente positivo de 1,813. El encontrar un producto óptimo, sin embargo, no es tan sencillo, puesto que existe el término de interacción $N_1 \cdot N_2$, con un coeficiente negativo (0,083) que crece con el producto del número de carbonos de los dos restos acilo, y que explica la reducción de actividad de los compuestos simétricos. Por ello, los

óptimos calculados a partir del modelo, representados en la Tabla 2.2, corresponden mayoritariamente a productos con un sustituyente largo y otro corto. En este caso, dado que el modelo es muy sencillo, y el número de derivados posibles es de 6^2 y se ha preparado un tercio del conjunto, varios óptimos han sido ya preparados en la serie de exploración, lo que no ocurre en el caso de series y modelos mas complejos.

Tabla 2.2. Algunos óptimos del modelo 2.5.



R	R'	Log (1/C) _{calc}	Log (1/C) _{exp}
H	Pent	2,553	2,55
H	But	2,397	2,46
Et.	Et	2,390	2,40
Me	Pent	2,388	-
Me	But	2,317	2,25
Pro	Pro	2,303	2,32
Me	Me	2,025	2,06
But	But	1,844	-
Pent	Pent	1,391	-

2.4 Alcance y limitaciones del análisis de Free-Wilson.

Usos del análisis de Free-Wilson en diversos campos y en combinación con otras técnicas QSAR se han revisado recientemente (Kubinyi, 1990), mostrando la flexibilidad y utilidad de la técnica. Especialmente en las primeras fases de la optimización de un prototipo, la técnica produce con gran facilidad buenos resultados, mostrando las diferentes contribuciones de los fragmentos utilizados, cuando hay términos de interacción no lineal, y abriendo en general la vía para la elaboración de modelos paramétricos mas sofisticados, todo ello siempre que se haya partido de una serie de exploración bien diseñada.

La principal limitación de la técnica radica, en que al establecer relaciones entre la actividad y la contribución de determinados grupos funcionales, no da ninguna información acerca de contribuciones de otros grupos no estudiados -a diferencia de los modelos paramétricos, que establecen relaciones entre la actividad y la contribución de ciertas propiedades fisico-químicas de los grupos funcionales-, por ello lo que se obtiene

como óptimo es la mejor combinación entre los grupos funcionales estudiados en la serie de exploración.

3.0 MODELOS PARAMETRICOS.

3.1 La metodología extratermodinámica (Fujita, 1990).

Los fármacos producen sus efectos en los sistemas biológicos a través de procesos complejos, en los cuales intervienen fenómenos de absorción, distribución en diferentes compartimentos, interacción con biomoléculas etc... Los tratamientos prácticos del problema deben evitar la necesidad de efectuar una descripción microscópica de cada uno de los pasos, si quieren mantenerse dentro de una simplicidad razonable. Uno de los primeros tratamientos que cumplía este requerimiento fue el desarrollado por C. Hansch y su equipo, al que hoy suele hacerse referencia con el nombre de metodología extratermodinámica (ME) ó método de Hansch.

La ecuación fundamental de la ME.

La acción del fármaco puede describirse como una serie de procesos encadenados (Hansch, 1973a) como se indica en la Fig. 3.1.

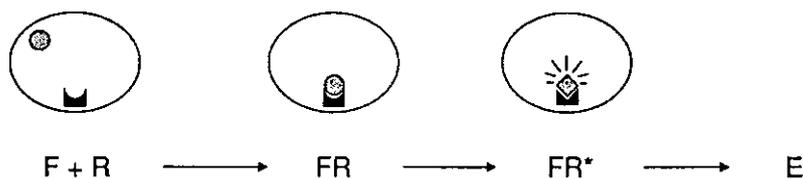


Fig. 3.1

Se asume que la velocidad del proceso es controlada por el paso III, el cual puede representarse más detalladamente como en la Fig. 3.2.

El fármaco (F) se une al receptor (R) para formar un complejo activado (FR^*), responsable del efecto (E). K_g es la constante de equilibrio del proceso y engloba constantes de equilibrio de otros posibles procesos intermedios. A partir del equilibrio se obtiene que

$$[FR^*] = K_g [F][R] \quad (\text{Ec. 3.1})$$

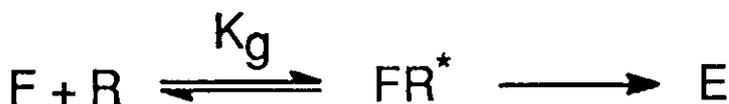


Fig. 3.2

como $[R] \gg [F]$, su concentración permanece aproximadamente constante, por lo cual puede incluirse en K_G , quedando la ecuación anterior en la forma

$$[FR^*] = K_G[F] \quad (\text{Ec. 3.2})$$

donde
$$K_G = K_b [R] \quad (\text{Ec. 3.3})$$

Por otra parte, la actividad ó potencia del fármaco suele cuantificarse en términos de dosis efectiva cincuenta (DE_{50}) o dosis equipotentes. Un fármaco es tanto más activo cuanto menores son estas dosis y por tanto

$$A = 1/[F] \quad (\text{Ec. 3.4})$$

y substituyendo $[F]$ por su valor en la ecuación 3.2

$$A = K_G/[FR^*] \quad (\text{Ec. 3.5})$$

Para una determinada dosis de fármaco administrada, $[FR^*]$ es una constante, por lo cual

$$A = K_G \text{ cte} \quad (\text{Ec. 3.6})$$

lo cual suele expresarse más comúnmente en forma logarítmica

$$\ln A = \ln K_G + \text{cte} \quad (\text{Ec. 3.7})$$

Así pues, el cambio de energía libre que se produce en la interacción F-R, se obtiene directamente como

$$\Delta G = -RT \ln K_G \quad (\text{Ec. 3.8})$$

Un postulado básico de la ME es que ΔG puede ser descompuesto en varios términos, cada uno de los cuales tiene en cuenta la contribución al cambio de energía libre que producen diferentes tipos de interacciones

fármaco-receptor (hidrofóbicas, electrónicas y estéricas). Esto puede expresarse como

$$-RT \ln K_G = \Delta G_h + \Delta G_e + \Delta G_s \quad (\text{Ec. 3.9})$$

o también

$$\ln K_G = f_h(x_h) + f_e(x_e) + f_s(x_s) \quad (\text{Ec. 3.10})$$

donde las funciones $f_h(x_h)$, $f_e(x_e)$ y $f_s(x_s)$ definen el modo en el que las características fisico-químicas del fármaco, cuantificadas por los parámetros x_h , x_e , x_s , influyen en su interacción con el receptor.

Por último, combinando la ecuaciones 3.7 y 3.10 se obtiene la ecuación básica de la ME,

$$\ln A = f_h(x_h) + f_e(x_e) + f_s(x_s) + \text{cte} \quad (\text{Ec. 3.11})$$

que establece la existencia de una relación entre la actividad de un fármaco y sus características fisico-químicas.

El término *extratermodinámica* que se aplica a esta metodología expresa el hecho de que las relaciones se describen en términos termodinámicos (como ΔG), aunque no se deducen de sus leyes (Martin, 1978).

El paradigma de Hansch.

Las asunciones fundamentales de la ME pueden formalizarse en una serie de puntos, la mayoría de los cuales se extraen de la ecuación fundamental (Unger, 1980):

1. La actividad biológica es función de la estructura del fármaco.
2. La estructura del fármaco implica ciertas propiedades globales como la hidrofobicidad, carga neta, solubilidad etc... y ciertas propiedades locales como la distribución de la hidrofobicidad, carga y volumen en determinadas posiciones de la molécula.
3. Estas propiedades globales y locales pueden ser cuantificadas mediante ciertos parámetros.
4. Siempre existe una función que relaciona los cambios en la actividad biológica con los cambios en las propiedades locales y globales, aunque puede que no sea ni sencilla ni evidente.

Modelos paramétricos.

Las funciones estructura-actividad que propone la ME constituyen un método mediante el cual se pueden buscar los productos más activos de entre un conjunto de candidatos. Habitualmente, se utiliza para buscar los óptimos de actividad de entre los posibles derivados (serie de exploración) de una molécula prototipo ó cabeza de serie, puesto que si se estudiaran moléculas estructuralmente muy distintas, raramente estarían implicados mecanismos de acción similares.

La ME postula la existencia de una función que relaciona la actividad (A) de un fármaco con sus propiedades fisico-químicas (x_h , x_e , x_s), pero no proporciona ningún método para calcularla. Para ello se han venido utilizado técnicas estadísticas de modelado empírico, que requieren llevar a cabo un proceso, en el que se distinguen las siguientes fases, ya indicadas en la Fig. 1.1:

1. Prototipo cabeza de serie.
2. Diseño de una serie de exploración de análogos del prototipo. Síntesis de los productos.
3. Caracterización fisico-química (x_h , x_e , x_s) y biológica (A) de cada uno de los productos de la serie.
4. Búsqueda mediante análisis de regresión de un modelo que relacione las constantes fisico-químicas y la actividad ($A = f(x_h, x_e, x_s)$).

3.2 Parámetros.

Para caracterizar en una molécula las propiedades fisico-químicas que aparecen en la ecuación fundamental de la ME, se utilizan lo que se denominan *parámetros extratermodinámicos*, variables que representan características fisico-químicas globales o locales de la molécula que influyen, desde un punto de vista energético, en la interacción de dicha molécula con el receptor. Estos parámetros se obtienen mediante distintos métodos, en general cuidadosamente diseñados para que representen características puras (hidrofóbicas, estéricas ó electrónicas) de la molécula o del sustituyente considerado. Sin embargo, raramente puede decirse que un parámetro represente solamente una característica de la molécula, en parte porque ciertas propiedades de las moléculas suelen presentarse asociadas, en mayor o menor medida.

La búsqueda de parámetros que representen propiedades puras, adecuados para relaciones estructura-actividad ha sido objeto de intensa investigación en los últimos años. Hoy en día, a la hora de caracterizar las

moléculas es posible escoger de entre la multitud de parámetros descritos en la literatura, cada uno de los cuales presenta ventajas e inconvenientes en su utilización (Waterbeemd, 1987).

Una clasificación que debe establecerse entre los parámetros es la que separa los referidos a propiedades de la molécula completa -parámetros moleculares- de los referidos a la contribución a una propiedad de un sustituyente o fragmento -parámetros fragmentarios-.

Parámetros moleculares.

Estos parámetros generalmente se obtienen mediante algún ensayo experimental que debe llevarse a cabo sobre cada una de los productos de la serie de exploración. Suelen representar propiedades globales de la molécula. En la Tabla 3.1 se han incluido algunos ejemplos:

Tabla 3.1. Algunos parámetros moleculares (Waterbeemd, 1987).

Símbolo	Descripción	Tipo ¹	Referencia
$\log P$	logaritmo del coeficiente de partición.	H	Hansch, 1979.
MR	refractividad molar molecular.	Es	Hansch, 1973b.
pK_a	constante de ionización.	El	Martin, 1978.
μ	momento dipolar.	El	Lien, 1982.
δ	desplazamiento químico (RMN).	El	Ciajolo, 1983.
ν	frecuencias (IR).	El	Lebedev, 1982.
K	constantes de equilibrio ó cinéticas.	.- ²	-

¹H = hidrofóbico, El = electrónico, Es = estérico; ² Depende del proceso estudiado.

Este tipo de parámetros tiene el inconveniente de que es necesario obtener las moléculas de la serie de exploración, y medirlos sobre cada una de ellas, por lo que no son adecuados en la fase de diseño experimental. Por otra parte, las metodías pueden ser laboriosas y no siempre es posible llevarlas a cabo.

Posiblemente, el parámetro molecular más utilizado es el logaritmo del coeficiente de partición ($\log P$), que expresa la hidrofobicidad global de la molécula. Suele determinarse mediante disolución del producto en un medio bifásico, habitualmente agua/*n*-octanol, agitación del sistema hasta alcanzar la situación de equilibrio y medida de la concentración de soluto en cada una de las fases. Este método, pese a ser el más aceptado, es muy laborioso. Para evitarlo se han propuesto multitud de procedimientos alternativos (Leo, 1971; Martin, 1978; Purcell, 1973; Rekker, 1979), incluyendo métodos computacionales, como los programas PRO-LOGP (Darvas, 1987) ó CHEMICALC-2 (Suzuki, 1990; 1991) que permiten un cálculo aproximado del coeficiente de reparto sin necesidad de medidas experimentales.

Parámetros fragmentarios.

Habitualmente la parametrización suele efectuarse sobre un conjunto de derivados de un producto cabeza de serie -la serie de exploración- en el que solo varía la sustitución en una/varias posiciones de la molécula. Entonces, es posible caracterizar los miembros de dicha serie mediante parámetros que hacen referencia únicamente a los sustituyentes variables.

De este modo pueden caracterizarse propiedades locales de la molécula, pero también propiedades globales, si se efectúa la aproximación de considerar que cada parte de la molécula hace una contribución aditiva al valor de la molécula completa. La utilización de parámetros fragmentarios hace innecesaria la síntesis previa de los productos a caracterizar, por lo cual es muy adecuada en la fase de diseño experimental.

Los parámetros fragmentarios suelen obtenerse a partir de medidas experimentales, que se llevan a cabo sobre moléculas preparadas por la unión de un núcleo estándar con el sustituyente y se escalan frente al valor del núcleo estándar sin sustituir. Así por ejemplo, el parámetro hidrofóbico de sustituyente π_x (Fujita, 1964) se obtiene

$$\pi_x = \log P_{RX} - \log P_{RH} \quad (\text{Ec. 3.12})$$

a partir de los valores de $\log P$ de un núcleo estándar R (generalmente benceno) substituido con X ($\log P_{RX}$) y sin substituir ($\log P_{RH}$).

Otra aproximación consiste en medir un parámetro experimental para una extensa serie de moléculas, considerar los fragmentos que los constituyen y someter los datos a un análisis estadístico que revele como contribuye cada uno de estos fragmentos al valor obtenido para la molécula completa. Un

ejemplo de este tipo de tratamiento son las constantes fragmentarias de Rekker (Nys, 1973; Rekker, 1977) o las de Hansch-Leo (Leo, 1975).

En la tabla 3.2 se han incluido algunos ejemplos de parámetros fragmentarios. Los valores, para la mayoría de los sustituyentes habituales en química orgánica, pueden encontrarse tabulados en la literatura (Hansch, 1979; Waterbeemd, 1987).

Tabla 3.2. Algunos ejemplos de parámetros fragmentarios (Waterbeemd, 1987).

Símbolo	Nombre, descripción	Tipo ¹	Referencia
π	constante hidrofóbica de sustituyente.	H	Fujita, 1964.
f	constante hidrofóbica fragmentaria.	H	Nys, 1973.
f, Fr	constante hidrofóbica fragmentaria.	H	Leo, 1975.
σ	constante de Hammett.	El	Hammett, 1940.
\mathcal{F}, \mathcal{R}	constantes de efecto inductivo y resonante.	El	Swain, 1968.
F, R	constantes posicionales de efecto inductivo y resonante.	El	Williams, 1976.
E_s	parámetro estérico de Taft.	Es	Taft, 1956.
MR	refractividad molar de sustituyente.	Es	Hansch, 1973b.
$L, B1, \dots, B5$	parámetros STERIMOL.	Es	Verloop, 1976.

¹H = hidrofóbico, El = electrónico, Es = estérico.

Otros tipos de parámetros.

No todos los parámetros tienen un origen experimental. También se pueden obtener a partir de :

1. Modelado molecular: parámetros STERIMOL (Verloop, 1976; Verloop, 1987).
2. Cálculos mecano-cuánticos: energía de HOMO y LUMO etc...

3. Estructura molecular: índices de conectividad (Kier, 1976; Kier 1983), peso molecular etc..

Mención aparte merecen las variables indicadoras, de tipo binario, que permiten combinar en una sola ecuación modelos obtenidos para series de moléculas ligeramente diferentes (Franke, 1984).

3.3 Diseño de la serie de exploración.

El diseño de la serie de exploración es la primera etapa de cualquier estudio de relaciones estructura-actividad. En este momento:

1. Se define el alcance de la variación estructural en los productos.
2. Se seleccionan los parámetros que caracterizarán los productos.
3. Se elige cada una de las moléculas a incluir en la serie de exploración.

La etapa de diseño es la que determina la *cantidad de información* que se introduce en el estudio. Es por ello que el éxito o fracaso de todo el proceso depende en gran medida de esta etapa. En ocasiones, es muy tentador intentar obtener modelos estructura-actividad a partir de series preexistentes, no diseñadas racionalmente. En estos casos, el éxito está condicionado a que la serie -de manera fortuita- contenga información suficiente, lo que no es habitual ni siquiera para series muy extensas.

Al plantearse el diseño de una serie experimental aparece la siguiente paradoja: para diseñar una serie experimental es necesario conocer el modelo, no obstante, el modelo no se conoce hasta que no se analiza la serie. La única posibilidad razonable es un procedimiento iterativo como el que se ilustra en la figura 1.2. Se diseña la serie partiendo de una hipótesis acerca del modelo, obteniéndose un modelo que puede ó no ajustarse a las expectativas iniciales, pero que produce información útil para mejorar el diseño en una nueva etapa de investigación.

Variación estructural.

Los estudios estructura-actividad suelen llevarse a cabo sobre un conjunto de derivados del prototipo cabeza de serie. En esta etapa se definen las posiciones de la molécula sobre las que se va a llevar a cabo la variación estructural, las cuales se eligen en función (a) de la influencia de los sustituyentes en estas posiciones sobre la actividad y (b) la accesibilidad sintética de los derivados.

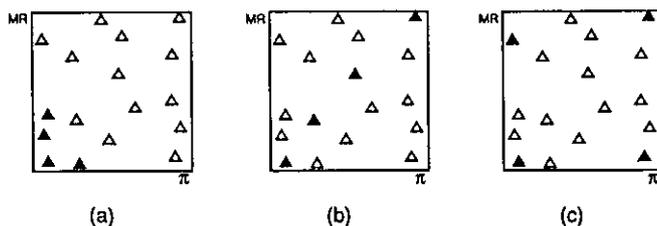


Fig. 3.3

Es importante explicitar la variación estructural emprendida. El conjunto de todos los posibles derivados del prototipo de partida constituye lo que se denomina *espacio experimental*. Cualquier modelo de relaciones estructura-actividad que se obtenga solo tiene validez, como máximo, dentro del espacio experimental utilizado y si se obtiene un óptimo será el producto más activo de este espacio experimental.

Criterios de selección de parámetros.

No existe ningún criterio simple ni único, para seleccionar qué parámetros deben utilizarse en un estudio de relaciones estructura-actividad. Ante un problema nuevo se debe comenzar por buscar en la literatura trabajos previos efectuados sobre el mismo tipo de productos y de no encontrarse, sobre productos parecidos o con la misma actividad etc... En general, en esta primera etapa, cualquier información puede ser orientativa respecto a las características físico-químicas que más intervienen sobre la actividad y los parámetros más adecuados para describirlas.

Es frecuente, no obstante, encontrar situaciones en las que no existe ninguna información previa. En estos casos, la única posibilidad es un compromiso inteligente; seleccionar unos pocos parámetros, cada uno representativo de una característica físico-química diferente y de carácter lo más general posible. Por ejemplo, es recomendable un análisis de Hansch clásico, utilizando un parámetro hidrofóbico (p.ej. π de Hansch), un parámetro estérico (p.ej. MR) y otro electrónico (p.ej. σ de Hammett). La serie diseñada conforme a estos parámetros es lo suficientemente informativa como para indicar (a) qué tipo de parámetros son significativos y (b) que tipo de dependencia existe, y así orientar sucesivas etapas de la investigación.

Elección de la serie.

El modelo se obtiene a partir de un pequeño conjunto de productos (la serie de exploración). Si se pretende que el modelo sea aplicable a todo el espacio experimental, la serie debe ser representativa de dicho espacio. Así pues, deberá ser suficientemente *informativa* para que las conclusiones que se deriven del modelo puedan aplicarse fuera de la propia serie.

La calidad de una serie de exploración viene definida principalmente por dos aspectos (Pleis, 1990; Pastor, 1991):

1. *Disimilaridad de la serie.* Los productos de la serie deben ser, respecto de los parámetros elegidos, diferentes entre sí. No tiene sentido intentar estudiar cómo influye una característica sobre la actividad si solo se estudian productos similares respecto a esta característica. Además, puesto que la serie debe ser una muestra del espacio experimental, el rango de variación de las características dentro de dicha serie debe ser equiparable al que existe en el espacio experimental.
2. *Ortogonalidad de la serie.* Los productos deben escogerse de tal forma que la variación en las características se produzca de manera independiente. Si la variación de una propiedad x_1 viene siempre acompañada de la variación en otra propiedad x_2 , no es posible conocer si el cambio de actividad de unos a otros productos se debe al cambio en la propiedad x_1 ó en la propiedad x_2 . En estas situaciones se dice que las propiedades x_1 y x_2 se encuentran *correlacionadas* en la serie. Una serie en la cual las correlaciones entre los parámetros son mínimas se dice que es *ortogonal*.

En diseño es frecuente ilustrar estas cualidades mediante gráficos bi o tridimensionales como los de la figura 3.3, en los que se sitúan los productos según los valores de sus parámetros, cada uno representado por un eje. En los gráficos de la figura 3.3 cada triángulo vacío indica un posible derivado del cabeza de serie de manera que el conjunto de todos ellos representa al espacio experimental. Los triángulos llenos señalan productos seleccionados para la serie de exploración.

La figura 3.3a muestra el ejemplo de una serie constituida por productos muy parecidos entre sí (serie poco disimilar). En la figura 3.3b se puede observar cómo la variación de ambos parámetros se produce de modo simultáneo: un valor alto de π viene siempre acompañado por un valor alto de MR (serie poco ortogonal). La serie representada en la figura 3.3c constituye un diseño correcto (serie disimilar y ortogonal).

Para lograr una selección correcta de los productos de la serie, una de las aproximaciones posibles es la aplicación de técnicas clásicas de diseño experimental, tales como el diseño factorial, el análisis de superficie de respuesta y técnicas de diseño óptimo (Box, 1971), ya sea de forma directa o adaptándolas a las características del problema. El lector interesado puede encontrar una amplia discusión de estos y otros métodos en una revisión eficiente (Pleiss, 1990).

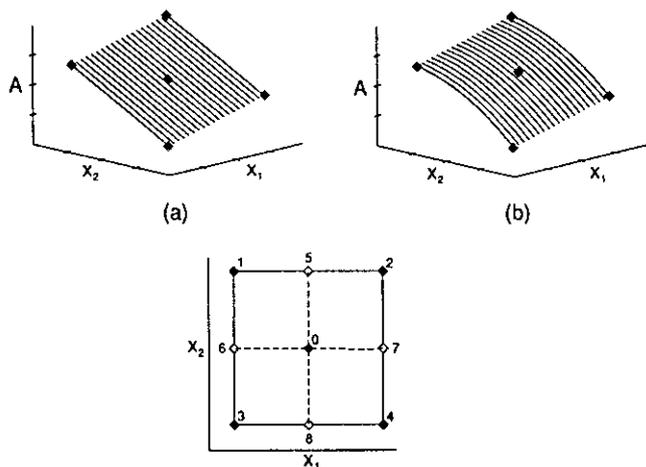


Fig. 3.4

Una cuestión previa a la elección del método de diseño es prever el tipo de modelo que se espera obtener -efectuar una hipótesis sobre el modelo-. En el área que nos ocupa suelen obtenerse modelos que pueden contener términos (a) lineales, $a_i x_i$; (b) rectangulares, $a_i x_i x_j$ y (c) cuadrados, $a_i x_i^2$. De entre ellos, los más frecuentes con diferencia son los lineales; los cuadrados solo suelen aparecer en parámetros hidrofóbicos -debido a la compleja influencia de la hidrofobicidad sobre la actividad del fármaco- mientras que los términos rectangulares son infrecuentes. Lo ideal sería diseñar series de exploración que permitieran el estudio de cualquier tipo de modelo, pero esto conduciría a series muy extensas. Una aproximación mucho más racional consiste en preparar, en un primer momento, una serie adecuada para un tipo de modelo frecuente. Si a partir de su análisis se sospecha un modelo más complejo, pueden añadirse nuevos productos a la serie inicial.

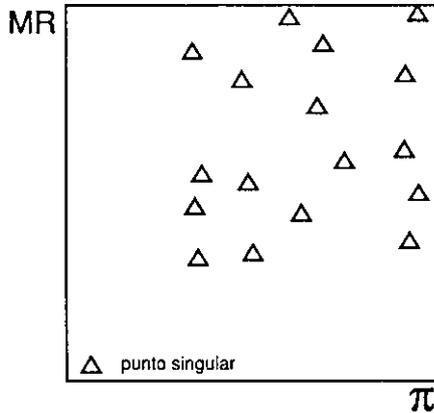


Fig. 3.5

En la Figura 3.4 se representa una situación hipotética en la que la actividad (A) es función de dos parámetros (x_1 y x_2). Si se sospecha un modelo compuesto solo por términos lineales ($A = a_1x_1 + a_2x_2 + b$, Fig. 3.4a), bastaría con un diseño factorial (puntos 1 a 4) o incluso factorial fraccionario (puntos 1 y 2, ó puntos 3 y 4). Este diseño no permite detectar la presencia de términos cuadrados ($A = a_1x_1 + a_2x_2 + a_3x_2^2 + b$, Fig. 3.4b) por lo que es conveniente introducir al menos un punto central (punto 0).

Si tras analizar los datos se sospecha la existencia de términos cuadrados, el punto central no sería suficiente para caracterizar el modelo y deben estudiarse series más completas, por ejemplo un diseño compuesto central (puntos 0 a 8). Los términos rectangulares pueden ser estudiados a partir de series obtenidas mediante diseños factoriales completos.

Otro aspecto importante del diseño es la elección los productos candidatos a ser incluidos en la serie. Es frecuente que entre los productos incluidos en el espacio experimental se encuentren algunos *puntos singulares*, productos de características tan especiales que son responsables de gran parte de la varianza de un parámetro en el espacio experimental (Figura 3.5). Suelen tener valores extremos para alguno/varios de los parámetros y como consecuencia tienden a ser incluidos en las series experimentales.

Estos productos, que suelen ser derivados con grupos cargados, o muy voluminosos etc... tienen frecuentemente comportamientos biológicos atípicos -no son absorbidos, son tóxicos, inactivos, etc.- por lo cual no es recomendable incluirlos en la serie de exploración. Se han propuesto diferentes métodos para evitar su influencia negativa en el diseño (Orozco, 1989; Pastor, 1991).

Los autores, conscientes de la importancia del diseño de series y de su dificultad han desarrollado los programas EDISFAR (Pastor, 1991) para asistir al investigador en esta tarea. Los programas EDISFAR se utilizan sobre ordenadores personales (PC-compatibles) y ofrecen, en un entorno de apariencia estándar (manejo con ratón, menús de persianas, datos en ventanas, etc...) el acceso a métodos de diseño especialmente adaptados a la generación de series de moléculas. Incluyen métodos de diseño D-óptimo (Mitchell, 1974) junto con métodos factoriales, factoriales fraccionales y compuestos, utilizando técnicas basadas en la metodología de Austel (Austel 1982a; 1982b; 1983). Evidentemente, la utilización de un programa no sustituye el conocimiento de los métodos que se utilizan, pero facilita el trabajo al evitar los cálculos y presentar al instante los gráficos e índices estadísticos que, de otro modo, hubieran requerido ser procesados mediante otros programas.

Ejemplo práctico.

El objetivo de este estudio de simulación es obtener derivados de 2-fenil-oxazolo[4,5]piridina I (Figura 3.6) que presenten actividad antiinflamatoria óptima (Pastor, 1991).

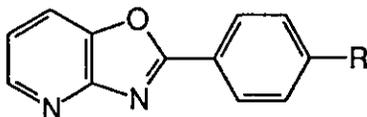


Fig. 3.6

La serie consta de derivados I con diferentes sustituyentes R en la posición *para* del fenilo. Se parametrizan con dos parámetros de sustituyente clásicos; π de Hansch (que describe hidrofobicidad) y σ de Hammett (que describe características electrónicas). En un estudio real la elección de los descriptores debería ser justificada, bien por estudios previos, bien como una tentativa inicial, como ya se expuso anteriormente.

Los sustituyentes candidatos a ser incluidos en el diseño se toman de la base de sustituyentes de EDISFAR; MONOBASE. En total hay 181 sustituyentes parametrizados, que se han representado en la Figura 3.7, con respecto a sus valores de π y σ . Siguiendo una aproximación sencilla, se considera el estudio de un modelo lineal del tipo

$$A = a_1\pi + a_2\sigma + b \quad (\text{Ec. 3.13})$$

para el cual un diseño factorial del tipo 2^2 es idóneo. Se incluye un punto central para comprobar si el modelo lineal es correcto (Box, 1971).

Una vez tomadas estas decisiones el resto del trabajo es realizado por EDISFAR. El usuario solo debe introducir el nombre de la base de parámetros -MONOBASE- y el modelo. Tomando las opciones por defecto el programa lleva a cabo un diseño factorial, de la siguiente manera:

1. Toma a partir de MONOBASE, de todos los sustituyentes para los que estén descritos los parámetros que utiliza el modelo (π y σ).
2. Normaliza entre 0 y 1 las escalas de los parámetros (Figura 3.7a).
3. Sitúa unos puntos marcadores (PM) en el espacio, que representan posiciones ideales para la serie (Figura 3.7b).
4. Presenta al usuario un listado con los substituyentes mas próximos a estos puntos marcadores, utilizando como criterio la distancia euclidiana normalizada (DEN) (Tabla 3.3).

Tabla 3.3. Diseño propuesto por EDISFAR

PM	π	σ	DEN	PM	π	σ	DEN
0	0	0		3	-	+	
SH			0.01	1-tetrazolil			0.04
NCCl ₂			0.01	PO(OCH ₃)			0.04
NNN			0.01	SOC ₂ H ₅			0.04
2-piridil			0.02	5-Cl-1-tetrazolil			0.05
CH ₂ Cl			0.04	COCH ₃			0.06
1	-	-		4	+	+	
NHCONH ₂			0.06	CF ₂ CF ₃			0.03
OH			0.06	SCF ₃			0.06
NHOH			0.07	C=O(OC ₆ H ₅)			0.08
NHCOOCH ₃			0.11	NNC ₆ H ₅			0.10
OCH ₃			0.13	SF ₅			0.10
2	+	-					
OC ₄ H ₉			0.04				
CH ₂ Si(CH ₃) ₃			0.05				
C(CH ₃) ₃			0.05				
CH(CH ₃) ₂			0.08				
C ₄ H ₉			0.08				

Como se aprecia en la figura 3.7, el espacio experimental contiene puntos singulares. Para evitar el efecto negativo de estos puntos -generalmente puntos aislados correspondientes a substituyentes con propiedades inhabituales-, EDISFAR sitúa el punto central del diseño sobre el valor medio

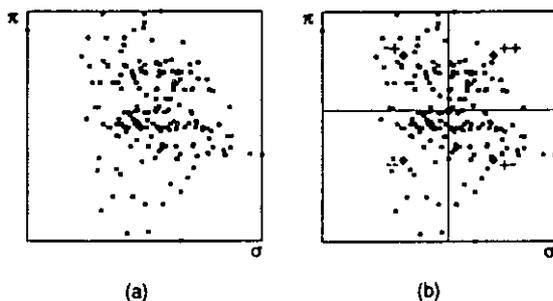


Fig. 3.7

de cada parámetro, y los puntos marcadores a una distancia del punto central que resulta de multiplicar un coeficiente (1,2 por defecto) por el valor de la desviación estándar del parámetro. De este modo el diseño se realiza correctamente sin necesidad de eliminar de la base los puntos singulares.

De entre los sustituyentes presentados por EDISFAR para cada punto marcador, el experimentador puede seleccionar para la serie exploratoria definitiva aquellos más accesibles sintéticamente, o que sospeche más estables etc... Las características de calidad de la serie seleccionada son presentadas de inmediato por EDISFAR (Tabla 3.4). Por este orden aparecen:

1. Varianza relativa media. Índice de disimilaridad que compara la dispersión -varianza- de la serie con la del espacio experimental. Un valor superior a 1 indica que la serie puede considerarse una muestra representativa del espacio experimental.
2. Matriz de correlación y su determinante. Son índices de ortogonalidad, el primero indica la correlación existente entre cada par de parámetros y el segundo es un índice global de la ortogonalidad de la serie. Una serie ortogonal debe presentar elementos de la matriz pequeños y un determinante próximo a 1.

3. Autovalores. Son también índices de ortogonalidad, sus valores deben ser parejos. Si los últimos autovalores son pequeños (explican menos del 5% de la varianza) la serie está demasiado correlacionada.

Tabla 3.4 Serie de exploración obtenida mediante EDISFAR.

PM	π	σ	sustituyente	DEN
0	0	0	SH	0.01
1	-	-	NHCONH ₂	0.06
2	+	-	OC ₄ H ₉	0.04
3	-	+	1-tetrazolil	0.04
4	+	+	CF ₂ CF ₃	0.03

Disimilaridad de la serie.

Varianza relativa media (VRM) = 1.4263

Ortogonalidad de la serie.

Matriz de correlación y autovalores.

	π	σ	Autovalor	%Var.	%Cum.
π	1.0000		1.0324	51.62	51.62
σ	0.0324	1.0000	0.9676	48.38	100.00

Determinante de la matriz de correlación (DMC) = 0.9989

3.4 Análisis de regresión.

Una vez diseñada la serie, se preparan los productos, se evalúa la actividad biológica de cada uno de ellos y se parametrizan según lo planeado. En este momento se dispone de dos tipos de información:

1. Información que debe ser explicada por el modelo -la actividad biológica-, que constituye la *variable dependiente*.
2. Información que explica -los parámetros-, que constituye el conjunto de *variables independientes*.

Ambos tipos de información se relacionan a través de una función matemática cuyo tipo se supone de antemano, puesto que el diseño se ha realizado considerando ese determinado tipo de función. En modelos estructura-actividad suelen encontrarse modelos lineales de la forma

$$A = a_1x_1+a_2x_2+\dots+a_nx_n+b \quad (\text{Ec. 3.14})$$

independientemente de que x_i pueda corresponder a un parámetro al cuadrado o a un producto de parámetros. Para calcular los valores de los coeficientes a_i y el término constante b lo más frecuente es utilizar técnicas de *análisis de regresión lineal múltiple* -generalmente abreviada como MLR- (Draper, 1981). No introduciremos aquí el fundamento del método, puesto que en la práctica este análisis se suele llevar a cabo mediante algún programa de ordenador. Mucho más importante es destacar la necesidad de comprobar la calidad del modelo obtenido.

Aparte de MLR también pueden utilizarse otras técnicas. El lector interesado puede encontrar una discusión general en (Waterbeemd, 1987) y un tratamiento bastante riguroso en (Draper, 1981).

Validación de modelos

Es casi una regla general que los datos que se obtienen mediante un ordenador suelen ser aceptados sin discusión, especialmente si no se conoce en profundidad el método implicado. No obstante, el modelo que se obtiene según el procedimiento descrito puede tener diferentes tipos de vicios y su calidad debe comprobarse.

El método de MLR proporciona los valores para los coeficientes que hacen que la función matemática propuesta explique al máximo como varían las actividades. Siempre existirá alguna variación no explicada -los residuales-, que corresponde a errores en la medida de los datos de actividad y a posibles efectos no incluidos en el modelo -parámetros no considerados, posibles efectos parabólicos, etc...-. La utilidad de un modelo depende de que la varianza explicada sea mucho mayor que la no explicada. En caso contrario podría ocurrir que el modelo no refleje ninguna relación real, sino que sea fruto púramente del azar. Esto puede contrastarse mediante el análisis de la varianza y test de F (Draper, 1981), y cuando el modelo no lo satisface se dice que es *no significativo*. Algo parecido ocurre para las variables, cuya intervención en un modelo explicando la actividad, puede ser significativa o no, contrastándose en este caso mediante un test de T. Encontrar variables no significativas ó incluso modelos completos no significativos no es preocupante, proporciona unas información valiosa indicando qué propiedades físico-químicas no intervienen en la actividad.

Esto nos permite, en próximas etapas de la investigación diseñar las series respecto a un menor número de parámetros ó respecto a parámetros diferentes.

Otro posible problema es que el modelo no sea *predictivo*, es decir que la función propuesta no se cumpla para productos no incluidos en la serie de exploración. También esto puede comprobarse mediante los adecuados análisis estadísticos, como por ejemplo mediante validación cruzada (*crossvalidation*) (Wold, 1991).

No todos los autores están de acuerdo respecto a qué es un modelo de buena calidad ni respecto a los criterios que estos deben cumplir (Mager, 1992). Lo que sí parece claro es que siempre deben establecerse unos criterios de validación y cumplirse, independientemente de que sean más o menos restrictivos .

Optimización

Para que el estudio de relaciones estructura-actividad se concluya con éxito debe obtenerse un modelo significativo y predictivo. Aun así, la validez del modelo estará siempre limitada al rango de parámetros explorado por la serie de exploración, fuera del cual nunca debe considerarse válido. De hecho, esta es la principal restricción a la hora de buscar el óptimo de actividad predicho por un modelo. Cuando el modelo contiene sólo términos lineales basta con buscar sustituciones que maximicen los términos con coeficientes positivos y que minimicen los términos con coeficientes negativos, siempre considerando que no se deben rebasar los valores mínimos y máximos de la serie experimental. Cuando los modelos son complejos y contienen términos rectángulos, cuadrados, etc... pueden utilizarse métodos automatizados (Takayama, 1983 ; Pastor, 1991). EDISFAR es capaz de realizar búsquedas automatizadas del óptimo, presentando un índice que indica el alejamiento del óptimo respecto del área explorada por la serie de exploración.

4.0 AGRADECIMIENTOS.

Los autores agradecen a la Comunidad de Madrid la concesión de una beca predoctoral (M.P.) y a la Comisión Interministerial de Investigación Científica y Técnica (CICYT) la financiación recibida (Proyecto PB90-0256). Los programas EDISFAR se distribuyen gratuitamente a cualquier institución sin ánimo de lucro que lo solicite.

5.0 BIBLIOGRAFIA.

- Austel, V. (1982) *Eur. J. Med. Chem.* 17, 9.
- Austel, V. (1982) *Eur. J. Med. Chem.* 17, 339.
- Austel, V. (1983) *Quant. Struct.-Act. Relat.* 2, 59.
- Austel, V. (1985) "QSAR and Strategies in the Design of Bioactive Compounds. *Proceedings of the Fifth European Symposium on Quantitative Structure-Activity Relationships*" J. K. Seydel Ed., VCH, Weinheim, p. 247.
- Bocek, K; Kopecky, J.; Krivukova, M.; Vlachova, D. (1964) 20, 667.
- Box, G. E. P.; Hunter, W. G.; Hunter, J. S. (1971) "Statistics for Experimenters", Wiley, New York.
- Bruice, T. C.; Kharash, N.; Winzler, R. J. (1956) *Arch. Biochem. Biophys.* 56, 305.
- Cativiela, C.; Elguero, J.; Mathieu, D.; Melendez, E.; Phan Tan Luu, R. (1983) *Eur. J. Med. Chem.-Chim. Ther.* 18, 344.
- Ciajolo, M. R.; Leij, F.; Tancredi, T.; Temussi, P. A.; Tuzi, A. (1983) *J. Med. Chem.* 6, 1060.
- Craig, P. N. (1984) *Drug Inform. J.* 18, 123.
- Darvas, F.; Erdos, I.; Teglas, G (1987) "QSAR in Drug Design and Toxicology", Elsevier. Amsterdam.
- Delgado, F.; Santiesteban, I.; Vega, J. A.; Minguez, J. M. (1993) *Resultados sin publicar*. El modelo corresponde a un ejercicio de Doctorado.
- Draper, N. R.; Smith, H.(1981) "Applied Regression Analysis", Wiley, New York.
- Franke, R. (1984) "Theoretical Drug Design Methods", Elsevier, Amsterdam.
- Free, S. M. y Wilson, J. W. (1964) *J. Med. Chem.* 7, 395.
- Fujita, T.; Iwasa, J.; Hansch, C. (1964) *J. Am. Chem. Soc.* 36, 1031.
- Fujita, T.; Ban, T. (1971) *J. Med. Chem.* 14, 148.
- Fujita, T.(1990) "The Extrathermodynamic Approach to Drug Design" en "Comprehensive Medicinal Chemistry" C. Hansch Editor. Pergamon Press, Oxford. Vol. 4, pag. 497.
- Hammett, L. P. (1940) "Physical Organic Chemistry", McGraw-Hill, New York.
- Hansch, C.; Fujita, T. (1964) *J. Am. Chem. Soc.* 86, 1616.
- Hansch, C. (1973a) "International Enciclopedia of Pharmacology and Therapeutics", Vol. I, Ed. Cavallito.
- Hansch, C.; Leo, A.; Unger, S. H.; Kim, K. H.; Nakaitani, D.; Lien, E.J. (1973b) *J. Med. Chem.* 16, 1207
- Hansch, C.; Leo, A. (1979) "Substituent Constants for Correlation Analisis in Chemistry and Biology", J. Wiley, New York.
- Hudson, D. R.; Bass, G. E.; Purcell, W. P. (1970) *J. Med. Chem.*13, 1184.
- Kier, L. B.; Hall, L. H. (1976) "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York.
- Kier, L. B. ;Hall, L. H. (1983) *J. Pharm. Sci.* 72, 1170.
- Kubinyi, H.; Kehrhahn, O.-H (1976) *J. Med. Chem.* 19, 578 .
- Kubinyi, H.; Kehrhahn, O.-H (1976) *J. Med. Chem.* 19, 1040.

- Kubinyi, H. (1990) "*The Free-Wilson Method and its Relationship to the Extratermodinamic Approach*" en "*Comprehensive Medicinal Chemistry*" C. Hansch Editor. Pergamon Press, Oxford. Vol. 4, pag. 589.
- Lebedev, R. S.; Lebedeva, I. R.; Lebedeva, L. A. (1982) *Khim. Farm. Zh.* 16, 1081.
- Leo, A.; Hansch, C.; Elkins, D. (1971) *Chem. Rev.* 71, 525.
- Leo, A.; Jow, P.C.Y.; Silipo, C.; Hansch, C. (1975) *J. Med. Chem.* 18, 865.
- Leo, A.; Hansch, C.; Jow, P. Y. C. (1976) *J. Med. Chem.* 19, 611.
- Lien, E. J.; Guo, Z. R.; Li, R. L.; Su, C. T. (1982) *J. Pharm. Sci.* 71, 641.
- Martin, Y. C. (1978) "*Quantitative Drug Design. A Critical Introduction*", Dekker, New York.
- Mitchell, T. J. (1974) *Technometrics.* 16, 203.
- Nys, G. G.; Rekker, R. F. (1973) *Eur. J. Med. Chem.* 8, 521.
- Orozco, M; Sanz, F. (1989) *An. Quim.* 85, 27.
- Pastor, M.; Alvarez-Builla, J. (1991) *Quant. Struct.-Act. Relat.* 10, 350.
- Pleiss, M. A.; Unger, S.H. (1990) "*Comprehensive Medicinal Chemistry*" Vol 4, 561, Pergamon Press, Oxford.
- Purcell, W. P.; Bass, G. E.; Clayton, J. M. (1973) "*Strategy of Drug Design; a Guide to Biological Activity*", Wiley, New York.
- Rekker, R. F. (1977) "*The Hidrophobic Fragmental Constant: Its Derivation and Application. A means of Characterizing Membrane Systems*", Elsevier. Amsterdam.
- Rekker, R. F. (1979) *Il Farmaco Ed. Sci.* 34, 346.
- Rekker, R. F. (1985) *Pharmacochem. Libr.* 8, 3.
- Suzuki, T.; Kudo, Y. (1990) *J. Comp.-Aided Mol. Design.* 4, 155.
- Suzuki, T. (1991) *J. Comp.-Aided Mol. Design.* 5, 149.
- Swain, C. G.; Lupton, E. C. (1968) *J. Am. Chem. Soc.* 90, 4328.
- Taft, R. W. (1956) "*Steric Effects in Organic Chemistry*", 556, Wiley, New York.
- Takayama, C.; Miyashita, Y.; Sasaki, S.; Yoshida, M (1983) *Quant. Struct.-Act. Relat.* 2, 121.
- Unger, S. H. (1980) "*Drug Design*", Vol IX. Academic Press, New York.
- Verloop, A.; Hoogenstraaten, W.; Tipker, J. (1976) "*Drug Design*", Vol VII. Academic Press, New York.
- Verloop, A. (1987) "*The STERIMOL Approach to Drug Design*", Dekker, New York.
- Waterbeemd, H. V.; Testa, B. (1987) "*Advances in Drug Research*", Vol 16, Academic Press, London.
- Williams, S. G.; Norrington, F. E. (1976) *J. Am. Chem. Soc.* 98, 508.
- Wold, S. (1991) *Quant. Struct.-Act. Relat.* 10, 191.