

Empleo de los índices topológicos y topográficos de Estrada en modelos QSAR de una familia de antibióticos γ -lactámicos

Ramón Carrasco Velar y Raúl Capote.

Centro de Química Farmacéutica, Avenida 21 y Calle 200, Reparto Atabey, Playa. Apartado Postal 16042, Código Postal 11600, Ciudad de La Habana, Cuba.

Recibido: 19 de septiembre de 2001. Aceptado: 15 de enero de 2002.

Palabras clave: QSAR, γ -lactámicos, índices topológicos, índices topográficos, regresión.
Key words: QSAR, γ -lactamic, topologic indices, topographic indices, regression.

RESUMEN. Los índices topológicos y topográficos son dos tipos de descriptores ampliamente utilizados en estudios de relación cuantitativa estructura-actividad (QSAR por sus siglas en inglés). Por otra parte, en el trabajo de búsqueda de modelos estadísticos adecuados para la modelación molecular se utiliza ampliamente el análisis de regresión múltiple (ARM). El procedimiento consiste en asociar un conjunto de descriptores estructurales o químico-físicos de las moléculas, con la actividad en un modelo de análisis de regresión múltiple estable que permita describir la muestra y, en el mejor de los casos, orientar la obtención de nuevos derivados según la tendencia que indique cada variable en la ecuación. El método de regresión por pasos es frecuentemente la técnica de elección para la eliminación de variables no significativas. A partir de una muestra reportada de 36 antibióticos γ -lactámicos sintetizados secuencialmente en una estrategia de diseño, se desarrolla un procedimiento iterativo de búsqueda de modelos predictivos de actividad utilizando ARM de forma similar al procedimiento aplicado por D.B. Boyd a esa familia, llamado *proyección iterativa de hipersuperficies* (PIH). Además de los descriptores empleados en el trabajo de referencia, se utilizaron índices topológicos y topográficos recientemente desarrollados por Estrada. Se comprueba que es posible emplear los nuevos índices topológicos y topográficos por su capacidad descriptora de la estructura química en estudios QSAR de antibióticos γ -lactámicos. El empleo del ARM de forma iterativa análoga al PIH, brinda resultados semejantes a este.

ABSTRACT. The topologic and topographic indexes are widely used descriptors in quantitative structure activity relationships (QSAR) due to the facilities of its calculations and interpretations. The Multiple Regression Analysis (MRA) is also a widely employed tool for molecular modeling. The procedure consists of associating a structural or physico-chemical descriptors set with the activity in a stable MRA model. This model must allow at least the description of the sample or, in other case, suggests the synthesis of new derivatives according to the tendency indicated by each variable in the equation. The forward stepwise regression is the preferred technique for the variable elimination. In this work it was started from a sample of 36 γ -lactamic antibiotics which were sequentially synthesized using a design strategy. An iterative procedure for the search of the best predictive activity model using MRA, was used. This procedure is similar to the *Hypersurface Iterative Projection* (HIP) method applied by D.B. Boyd to these compounds family. Topologic and topographic descriptors recently developed by Estrada were used together with the others reported in the original work. The possibilities of employing the indexes developed by Estrada in a QSAR study of γ -lactamic antibiotics, is demonstrated. The iterative use of MRA affords similar results to HIP.

INTRODUCCION

Desde el inicio de la época de los estudios de correlación estructura química-actividad biológica, ha sido incesante la búsqueda de los procedimientos estadísticos más adecuados para el procesamiento de los datos. Son diversos los problemas que se han presentado en este aspecto, entre los que se encuentran, la disminución del número de variables, la colinealidad entre ellas, la linealidad o no de los modelos y su carácter predictivo, criterios de selección de la muestra de estudio y de las variables a emplear, etc. Entre las técnicas más populares para la disminución del número de variables está el análisis de componentes principales entre otros análisis *clusters*; pero la técnica de procesamiento de datos multivariados más ampliamente extendida en los estudios de relación cuantitativa estructura-actividad (QSAR, Quantitative Structure Activity Relationships) es el análisis de regresión múltiple (ARM), y en él, se utiliza el método de regresión por pasos para la reducción de variables.

En cuanto a la colinealidad entre variables, las opiniones de los autores difieren bastante, aunque se ha llegado a un consenso en el que se acepta dentro de un modelo, una correlación entre ellas que oscile entre 0,6 y 0,7 como máximo,¹ aunque hay autores que aceptan valores mayores de 0,9.² En años recientes,^{3,4} salió a la luz un trabajo en el que se desarrolló un procedimiento de

ortogonalización de variables dentro del ARM al que se denominó análisis de componentes dominantes (ACD), que vino a poner una especie de punto final alrededor de esta controversia, que surgió desde la conocida ecuación parabólica de Corwin Hansch.⁵ En ese trabajo, se demuestra que variables con coeficiente de correlación mutuo hasta de 0,98 no pueden ser desestimados en una ecuación de regresión. Cuando dos variables tienen un $R \neq 1$ existe un contenido de información de una que no está contenido en la otra lo que da lugar a un residual. La eliminación de una de las dos conduciría a una pérdida de información estructural y se concluye que solamente cuando $R = 1$ debe eliminarse una. Se justifica, desde un punto de vista estadístico tanto el modelo de Hansch como el bilineal de Kubinyi⁶ los cuales responden, desde el punto de vista del modelo, al fenómeno que describen, independientemente de la socorrida colinealidad.

En cuanto al carácter predictivo de una ecuación, este dependerá en buena medida del criterio de selección de las muestras de entrenamiento y de prueba. Esta selección fijará *a priori* las condiciones por las que se regirá el estudio, y fijará los valores máximo y mínimo para interpolaciones. La extrapolación buscando compuestos más potentes, siempre será algo discutible. Es necesario para ello que el modelo describa realmente el fenómeno analizado y no el comportamiento de la muestra, y la simplificación de ellos constituye una aproximación adicional que puede alejarlo de la realidad. Es por eso que una tercera muestra de validación que incluya compuestos no vistos previamente puede ser importante para aumentar la fiabilidad del modelo. Este procedimiento, junto con la aplicación del análisis de validación cruzada desarrollado por Cramer III⁷ deben dar un margen de seguridad mayor en el carácter predictivo del modelo.

Pero no es solamente una buena selección de la muestra la que determina la calidad del modelo. Debe disponerse de descriptores adecuados tanto de la estructura química como de diversas propiedades químico-físicas de las moléculas si se desea lograr un resultado predictivo.

Recientemente, se introdujo y desarrolló un nuevo tipo de índices calculados a partir de la matriz de adyacencia de los vértices según el algoritmo de Randiæ.⁸ Estos son: los índices Ω (ponderada la matriz A por

los órdenes de enlace), Ω_q y Ω_q^c , (ponderada la matriz A por las cargas en ambos, y el último con una corrección espacial)⁹. También se definieron los índices ϵ (índice topológico calculado a partir de la matriz de adyacencia de las aristas del grafo químico E) y ϵ_p (índice topográfico calculado a partir de la matriz E ponderada por los órdenes de enlace).⁹

Ahora bien, en la búsqueda de un nuevo fármaco, generalmente se parte de una muestra inicial reducida de compuestos, los que por lo general, no son tan potentes como se desea, o de una muestra mayor con datos de la literatura, en la que aparecerán todos los problemas mencionados en párrafos anteriores. Esa muestra reducida puede ser, bien el resultado de un trabajo intuitivo del químico que por analogía estructural diseñó una serie de experimentación o serie cero, o resultado del aislamiento y purificación de compuestos activos naturales los que pueden haber experimentado transformaciones químicas. Este es uno de los estados de la investigación en las que el especialista en modelación molecular puede intervenir para brindar la explicación teórica de la capacidad de respuesta de los compuestos objeto de estudio y proponer las próximas modificaciones estructurales a realizar para aumentar potencia, biodisponibilidad o características farmacocinéticas.

El presente trabajo pretendió evaluar la capacidad de algunos índices topológicos y topográficos para la modelación molecular, así como comparar el método PIH con el ARM.

MATERIALES Y METODOS

El sistema bicíclico de la pirazolidinona (Fig. 1) que es un bioisómero de los anillos b-lactámicos^{10,11} fue el sustrato para este trabajo. En este sistema, el segundo nitrógeno es quien confiere la elevada reactividad del grupo acilo frente a nucleófilos. Un trabajo reciente de D.B. Boyd dio a conocer los resultados de un estudio QSAR empleando un procedimiento estadístico que denominó *proyección iterativa de hipersuperficies* (PIH) aplicado a esta familia.¹⁰

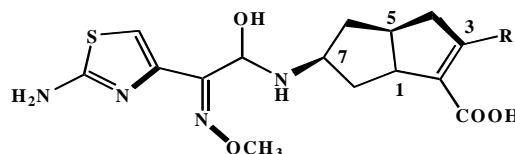


Fig.1. 7-amiotiazolmetiloxima acetamido pirazolidinona bicíclica, 3-sustituida.

Para el estudio QSAR se combinarán los índices topográficos de Estrada ${}^p\epsilon$ y ${}^p\Omega$, en los que (p) representa el grado del índice,¹² con los valores de σ_p (constante electrónica de sustituyentes en posición *para*), π (constante fragmental de lipofiliidad) y MR (refractividad molar) reportados para los diferentes sustituyentes en la molécula de γ -lactama.

El método proyección iterativa de hipersuperficies

Este método está basado en gráficos computacionales tridimensionales y tiene sus antecedentes en la graficación de Craig,¹ según la cual, la localización de los sustituyentes en un espacio de parámetros dado, se puede visualizar con un gráfico bidimensional. Cualquier par de parámetros de sustituyente es posible graficarlo a lo largo de los dos ejes. La selección de los pares se hace preferentemente si no son colineales. El objetivo es establecer comparaciones entre los diferentes sustituyentes y, atendiendo a semejanzas y diferencias, establecer una determinada estrategia de síntesis. A finales de los setenta, Cramer y col. emplearon esta forma de graficación con líneas de contorno para señalar las regiones del espacio de parámetros que brindaban la mejor actividad anti-alérgica en una serie de compuestos.² De hecho, este mapa de relieve añadía una tercera dimensión. Esta graficación se empleaba solamente al inicio del estudio y posteriormente, se continuaba con un análisis QSAR según el modelo clásico de regresión de Hansch. En el método PIH se realiza este mismo procedimiento, pero se continúa de forma iterativa añadiendo los nuevos compuestos una vez sintetizados y evaluados. Este método de graficación puede realizarse con las opciones gráficas del STATISTICA, para visualizar el gráfico tridimensional en forma de gráficos de contorno. En el trabajo de referencia,¹⁰ se partió de una primera muestra de 14 compuestos y mediante una graficación de Craig, se orientó hacia un nuevo espacio de los parámetros, para diseñar su segundo grupo de compuestos a sin-

tetizar y evaluar, y así sucesivamente hasta llegar a 36. Según este procedimiento, cada nuevo grupo de compuestos se fue integrando con el anterior para muestras de 19, 25 34 y 36 derivados.

En este trabajo se siguió la misma línea de pensamiento desarrollada por Boyd en cuanto a la *iteración* al emplear ARM, para lo cual se realizó el análisis de la muestra inicial de 14 compuestos, y se fue incrementando la muestra siguiendo la secuencia de síntesis y evaluación biológica dada por ellos.

Dado que el análisis se centró en el efecto del sustituyente en la posición 3 del sistema de anillos, los cálculos teóricos para la ponderación de los índices se realizaron solamente al fragmento correspondiente unido a un grupo vinilo. La tabla 1 muestra el grupo de sustituyentes empleados, los valores de las constantes empleadas y de log CMI (concentración mínima inhibitoria) reportados así como los valores calculados a los diferentes índices ${}^p\epsilon$ y ${}^p\Omega$ correspondientes.

Todas las estructuras fueron dibujadas con el programa HyperChem³ y optimizadas por el método MM⁺ incluido. La matriz Z de cada una de las geometrías resultantes se utilizó como dato de entrada para la optimización total de geometría por el método semiempírico AM1⁴ con una versión del sistema MOPAC 6.0^{5,6} programado para WINDOWS.⁷ A las estructuras se les realizó una optimización total de geometría utilizando el método BFGS.⁸ Se utilizaron las opciones de cálculo PRECISE y MMOK, y de salida VECTORS y BONDS. En todos los casos, se alcanzó la convergencia.

El procedimiento seguido para el análisis de los datos fue determinar la mejor ecuación de regresión para la muestra inicial de 14 derivados, predicción de los cinco siguientes, análisis del comportamiento de las variables en las muestras y posterior inclusión en la muestra para repetir el tratamiento estadístico con n = 19 compuestos y predicción de los próximos seis empleando el paquete STATISTICA.⁹ Este procedimiento se repitió iterativamente para las muestras restantes. Como indicadores de la condición de *outliers* se utilizaron: *el residual estándar*, *la distancia de Mahalanobis*, *los residuales borrados* y *la distancia de Cook*. Aquellos compuestos que resultaron *outlier* en tres o más de los criterios anteriores, se excluyeron de la muestra y se reinició el análisis.

RESULTADOS Y DISCUSION

En los estudios QSAR, se utiliza una forma común para reportar los resultados de actividad. Se le brinda al lector el conjunto completo de los datos, se discute y se proponen conclusiones. El lector también puede hacer sus propias interpretaciones, pero lo que no se obtiene nunca es el orden en que fueron sintetizados los compuestos que se reportan, ni las valoraciones que se hicieron en cada paso del proceso de modelación y diseño. Es por eso que el trabajo de Boyd¹⁰ brinda una facilidad exclusiva para modelar junto con ellos y poder corroborar o no nuevas hipótesis y enfoques, dado que describen las diferentes etapas que se vencieron para diseñar los compuestos que servirían como propuesta de síntesis y evaluación biológica en pasos posteriores a partir de una muestra inicial.

Tratamiento estadístico por métodos de regresión múltiple. Muestra de N = 14

En el trabajo se obtuvo un conjunto de modelos de regresión que involucraron diferentes variables descriptoras (Tabla 1).

Mediante análisis de regresión múltiple para la muestra inicial de 14 compuestos se llega a la ecuación I, como "mejor ecuación".

Todos los términos de la ecuación son significativos con valores de $p < 0,01$ para un 99,99 % de fiabilidad. Se encontró una correlación entre variables de $r = 0,60$.

Se conoce que los valores de σ_p se hacen mayores en la misma medida que aumenta el carácter electroceptor del grupo sustituyente. Además, se ha demostrado previamente¹ que los valores σ de Hammett correlacionan con la reactividad de las cefalosporinas, en las que existe una deslocalización análoga. Según la ecuación I, el peso principal en la actividad biológica está asociado a

σ_p . Valores mayores de σ_p deben mejorar la potencia de los compuestos al disminuir el log CMI. Grupos sustituyentes en la posición 3 con valores de ${}^2\Omega$ menores deben incrementar la potencia de los compuestos. Según los resultados teóricos alcanzados con el método PIH, se sintetizaron y evaluaron los compuestos del 15 al 19, acorde con el nuevo espacio de parámetros de sustituyentes al que debían orientarse para mejorar la potencia de los compuestos a sintetizar. A partir de la ec. I, que combina σ_p y ${}^2\Omega$ se ve que al comparar las medias de ambas muestras (Tabla 2), —la de 14 y 5 compuestos respectivamente— se cumple con la tendencia de comportamiento que sugiere dicha ecuación en la que se ha utilizado una variable (${}^2\Omega$) no incluida por Boyd en su trabajo. Esto representa a su vez, una coincidencia en la predicción de la actividad de compuestos nuevos por ambos procedimientos.

Tratamiento estadístico de la muestra de 19 compuestos

El análisis de regresión múltiple a esta muestra brindó en un primer paso la ecuación II como la "mejor", con una significación estadística para cada término $< 0,001$ al igual que la anterior. Después de eliminar como *outliers* estadísticos los compuestos 11 y 15 se obtiene la ecuación III, también con una significación estadística para cada término mayor del 99,99%. Esta ecuación mejora de manera notable el modelo, aunque el número de variables se incrementa y la relación casos/variables pasa de 8,5:1 a 4:1.

Si atendiendo a que entre las variables ${}^2\Omega$ y ${}^1\epsilon$ existe un coeficiente de correlación de 0,77 se elimina la última y rehacemos el análisis de regresión múltiple por pasos, se obtiene la ecuación IV con un nivel de significación igual a las anteriores. Este

$$\log CMI = -2,698(\pm 0,587)\sigma_p + 0,443(\pm 0,137){}^2\Omega + 1,746(\pm 0,234) \quad (I)$$

$$R = 0,812 \quad F_{(2,11)} = 10,682 \quad p < 0,002 \quad 64 \quad s = 0,400 \quad N = 14$$

$$\log CMI = -2,299(\pm 0,503)\sigma_p + 0,373(\pm 0,116){}^2\Omega + 1,728(\pm 0,262) \quad (II)$$

$$R = 0,778 \quad F_{(2,16)} = 12,259 \quad p < 0,000 \quad 59 \quad s = 0,475 \quad N = 19$$

$$\log CMI = -3,024(\pm 0,239)\sigma_p + 0,769(\pm 0,081){}^2\Omega + 0,222(\pm 0,044){}^1s$$

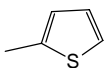
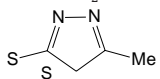
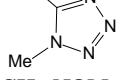
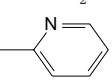
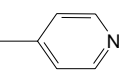
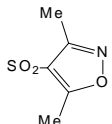
$$+ 0,132(\pm 0,040)\pi + 2,130(\pm 0,149) \quad (III)$$

$$R = 0,971 \quad F_{(4,12)} = 50,053 \quad p < 0,000 \quad 00 \quad s = 0,205 \quad N = 17$$

$$\log CMI = -2,908(\pm 0,397)\sigma_p + 0,480(\pm 0,088){}^2\Omega + 1,640(\pm 0,192) \quad (IV)$$

$$R = 0,903 \quad F_{(2,14)} = 31,084 \quad p < 0,000 \quad 01 \quad s = 0,342 \quad N = 17$$

Tabla 1. Pirazolidinonas empleadas en el estudio y variables que aparecen en los modelos de regresión.

No.	R	Log MIC	σ	π	MR	${}^2\Omega$	${}^1\Omega_{qC}$	${}^2\Omega_q$	${}^4\Omega$	${}^5\Omega_q$	${}^5\Omega_{qC}$	${}^1\varepsilon$	${}^2\varepsilon_p$	${}^4\varepsilon_p$
1	CO ₂ Me	1,050	0,45	-0,01	12,87	1,32	1,54	0,77	0,16	0,00	0,00	3,86	1,92	0,25
2	PO ₃ Me ₂	1,660	0,53	-1,18	21,87	2,99	2,15	1,33	0,71	0,00	0,00	5,08	3,89	1,11
3	H	2,320	0,00	0,00	1,03	0,00	0,46	0,00	0,00	0,00	0,00	0,00	0,00	0,00
4	CF ₃	1,870	0,54	0,88	5,02	2,87	1,29	0,79	0,00	0,00	0,00	3,23	2,36	0,00
5	PO ₃ Et ₂	2,290	0,60	-0,10	31,16	3,65	3,32	1,65	1,70	0,21	0,20	6,49	5,01	1,81
6	COMe	0,390	0,50	-0,55	11,18	1,13	1,38	0,81	0,00	0,00	0,00	2,99	1,5	0,00
7	COPh	1,080	0,43	1,05	30,33	2,29	2,98	1,92	0,90	0,40	0,38	6,77	3,69	1,30
8	Ph	2,410	-0,01	1,96	25,36	1,76	2,52	1,57	0,68	0,32	0,39	5,56	2,87	1,07
9		2,410	0,05	1,61	24,04	1,65	2,95	2,23	0,67	0,25	0,24	4,85	2,58	1,01
10	CH ₂ Oac	2,410	0,05	-0,17	16,48	1,91	2,02	1,12	0,39	0,09	0,08	4,41	2,48	0,50
11	CONHMe	1,990	0,36	-1,27	14,57	1,32	1,66	0,85	0,16	0,00	0,00	3,86	1,9	0,24
12	CO ₂ Pr	1,350	0,45	1,03	21,33	2,10	2,61	1,42	0,57	0,15	0,15	5,28	3,01	0,76
13	CO ₂ Et	1,260	0,45	0,51	17,47	1,73	2,12	0,98	0,48	0,06	0,06	4,57	2,5	0,63
14	CO ₂ H	1,780	0,00	-4,36	6,05	1,12	1,11	0,55	0,00	0,00	0,00	2,99	1,5	0,00
15	COCH ₂ F	1,660	0,70	-0,39	9,78	1,31	1,53	0,83	0,16	0,00	0,00	3,86	1,95	0,25
16	CONHPh	1,350	0,38	0,49	35,35	2,75	3,22	2,03	1,08	0,42	0,40	7,42	4,14	1,44
17	CO ₂ CH ₂ Ph	1,840	0,45	1,84	37,20	3,10	3,61	2,31	1,21	0,49	0,47	8,13	4,61	1,64
18	COEt	0,630	0,48	0,06	15,83	1,30	1,94	1,04	0,16	0,00	0,00	3,86	1,94	0,25
19	CN	-0,030	0,66	-0,57	6,33	0,39	0,89	0,33	0,00	0,00	0,00	2,71	0,91	0,00
20	SO ₂ Me	-0,390	0,72	-1,63	13,49	2,78	3,85	3,46	0,00	0,00	0,00	3,23	2,51	0,00
21	SCH ₃	2,260	0,00	0,61	13,82	0,72	2,36	1,23	0,00	0,00	0,00	2,71	1,14	0,00
22	SO ₂ Ph	0,030	0,68	0,27	33,20	3,51	5,12	4,54	1,16	0,94	0,89	7,09	4,71	1,49
23	SO ₂ Et	-0,210	0,77	-1,11	18,14	2,68	4,18	4,06	0,17	0,00	0,00	4,16	3,01	0,31
24	COCF ₃	2,410	0,80	0,02	11,17	3,35	1,76	1,13	0,36	0,00	0,00	4,46	3,19	0,36
25	SO ₂ N(Me) ₂	0,390	0,65	-0,78	21,88	3,44	4,15	4,55	0,30	0,00	0,00	4,52	3,67	0,37
26	NHCO ₂ Me	2,410	-0,17	-0,38	15,74	1,72	1,76	0,87	0,39	0,05	0,04	4,57	2,39	0,53
27		1,810	0,02	0,29	32,57	2,64	4,19	4,04	1,17	0,97	0,94	6,01	3,53	1,37
28		2,050	0,02	-0,50	26,66	2,59	3,28	2,41	1,18	0,39	0,37	6,06	3,58	1,44
29	CH=NOMe	1,870	0,30	0,40	15,73	1,13	1,59	0,68	0,25	0,05	0,04	4,12	1,88	0,43
30	CH=NOCH ₂ -CH ₂ Ph	1,810	0,30	2,49	46,40	3,26	4,15	2,54	1,42	0,57	0,54	9,10	5,06	2,05
31		2,110	0,09	0,50	23,03	1,77	2,39	1,43	0,69	0,26	0,25	5,56	2,88	1,08
32		2,200	0,18	0,46	23,03	1,77	2,38	1,45	0,69	0,25	0,24	5,56	2,87	1,08
33	CH=NOCH ₂ -CH ₂ =CH ₂	1,870	0,30	0,90	25,67	1,61	2,29	1,08	0,44	0,09	0,08	5,54	2,71	0,80
34	CH ₂ =CH ₂	2,380	-0,04	0,82	10,99	0,49	1,09	0,45	0,00	0,00	0,00	2,71	1	0,00
35		0,270	0,77	-0,48	34,03	4,42	5,39	4,82	2,28	0,84	0,80	7,38	5,31	1,92
36	SO ₂ CH ₂ CH ₂ -NHAc	0,270	0,77	-2,14	32,21	4,30	5,29	4,72	1,12	0,68	0,62	6,56	4,79	1,20
	Máximo	2,41	0,80	2,49	46,40	4,42	5,39	4,82	2,28	0,97	0,94	9,10	5,31	2,05
	Mínimo	-0,39	-0,17	-4,36	1,03	0,00	0,46	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Mediana	1,81	0,44	0,04	19,74	1,84	2,37	1,38	0,42	0,06	0,05	4,57	2,79	0,58
	Media	1,48	0,37	0,02	20,58	2,14	2,63	1,83	0,57	0,21	0,20	4,87	2,86	0,74

Los valores de log CMI son la media geométrica de las determinaciones frente a diez microorganismos diferentes, cuatro Gram (-) y seis Gram (+). Los datos microbiológicos fueron reportados en forma de log CMI del material racémico, tal como fue reportado en el trabajo original.

Tabla 2. Valor promedio de las variables en cada una de las muestras.

Ecuación 1												
n = 14		15-19		20-25		26-34		35-36				
Variabes	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p		
Promedio	1,847	0,314	1,770	0,534	2,744	0,603	1,885	0,111	4,358	0,770		
Ecuación 2												
n = 19		20-25		26-34		35-36						
Variabes	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p		
Promedio	1,827	0,372	2,744	0,603	1,885	0,111	4,358	0,770				
Ecuación 3												
n = 17				20-25				26-34				
Variabes	${}^2\Omega$	σ_p	${}^1\varepsilon$	π	${}^2\Omega$	σ_p	${}^1\varepsilon$	π	${}^2\Omega$	σ_p	${}^1\varepsilon$	π
Promedio	1,887	0,354	4,600	0,146	2,744	0,603	4,360	-0,437	1,885	0,111	5,469	0,553
35-36												
${}^2\Omega$	σ_p	${}^1\varepsilon$	π									
4,358	0,770	6,969	-1,310									
Ecuación 4												
n = 17		20-25		26-34		35-36						
Variabes	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p	${}^2\Omega$	σ_p		
Promedio	1,887	0,354	2,744	0,603	1,885	0,111	4,358	0,770				
Ecuación 5												
n = 24			26-34			35-36						
Variabes	${}^2\Omega$	σ_p	${}^2\Omega_q$	${}^2\Omega$	σ_p	${}^2\Omega_q$	${}^2\Omega$	σ_p	${}^2\Omega_q$	${}^2\Omega$	σ_p	
Promedio	2,698	0,416	1,693	2,744	0,603	1,661	4,358	0,770	4,772			
Ecuación 6												
n = 23			26-34			35-36						
Variabes	σ_p	${}^4\Omega$	${}^5\Omega_q$	σ_p	${}^4\Omega$	${}^5\Omega_q$	σ_p	${}^4\Omega$	${}^5\Omega_q$	σ_p	${}^4\Omega$	${}^5\Omega_q$
Promedio	0,400	0,456	0,144	0,111	0,691	0,290	0,770	1,701	0,761			
Ecuación 7												
n = 33					35-36							
Variabes	σ_p	${}^1\Omega_{qC}$	${}^2\Omega$	π	σ_p	${}^1\Omega_{qC}$	${}^2\Omega$	π	σ_p	${}^1\Omega_{qC}$	${}^2\Omega$	π
Promedio	0,330	2,486	1,963	0,096	0,770	5,340	4,358	-1,310				
Ecuación 8												
n = 32				35-36								
Variabes	σ_p	${}^4\varepsilon_p$	${}^5\Omega_{qC}$	σ_p	${}^4\varepsilon_p$	${}^5\Omega_{qC}$	σ_p	${}^4\varepsilon_p$	${}^5\Omega_{qC}$	σ_p	${}^4\varepsilon_p$	${}^5\Omega_{qC}$
Promedio	0,318	0,717	0,177	0,770	1,558	0,710						
Ecuación 9												
n = 32				35-36								
Variabes	σ_p	${}^4\varepsilon_p$	${}^5\Omega_{qC}$	σ_p	${}^4\varepsilon_p$	${}^5\Omega_{qC}$	σ_p	${}^4\varepsilon_p$	${}^5\Omega_{qC}$	σ_p	${}^4\varepsilon_p$	${}^5\Omega_{qC}$
Promedio	0,318	0,717	0,177	0,770	1,558	0,710						

resultado conduce de hecho a una ecuación con los mismos descriptores que en la ecuación I. Si se comparan los errores de los coeficientes de cada término en ambas (σ_p , ${}^2\Omega$ y la constante), se encuentra que coinciden los intervalos en todos los casos. Por otra parte, el resto de los estadígrafos mejora notablemente con respecto a la ecuación I con la inclusión de los nuevos compuestos. Es decir, sustituyentes que presenten valores de σ_p mayores y de ${}^2\Omega$

menores, deben incrementar la potencia de los compuestos. No obstante, debe pensarse que los aumentos de σ_p implican un incremento en la reactividad y por ende, no debe esperarse un aumento indefinido de la actividad biológica porque la reactividad tampoco aumenta indefinidamente. Otro tanto debe considerarse respecto al índice ${}^2\Omega$, pues el tamaño de los grupos sustituyentes y los órdenes de enlace dentro de él, guardan necesariamente una deter-

minada relación, no simple, con el carácter electrónico del índice. Un resultado similar se obtiene al analizar la ecuación II.

El índice topográfico ${}^2\Omega$ se calcula a partir de la matriz de adyacencia de los vértices del grafo químico ponderados por los órdenes de enlace. Una disminución de su valor se alcanza reduciendo el tamaño del sustituyente (menor número de vértices), el orden de enlace de sus átomos o ambos considerando frag-

mentos que incluyan dos vértices. Esto sugiere la disminución de sustituyentes con dobles enlaces o mayores y reducir la presencia de heteroátomos, algo que se contradice con la necesidad de grupos electroaceptores, los que por regla general contienen insaturación y heteroátomos, por lo tanto, el aumento del carácter electroceptor de los grupos sustituyentes no debe lograrse a partir de aumentar su conjugación, etc., sino de su electronegatividad.

Por otra parte, según la ecuación III, ϵ_p y ${}^1\epsilon$ deben aumentar en tanto ${}^2\Omega$ y π deben disminuir para mejorar la potencia. Con la muestra de 20 a 25 así ocurre para σ_p y π , pero no con ${}^2\sigma$ y ${}^1\epsilon$. No obstante, se logra un aumento promedio de la actividad predicha. La contribución media de cada parámetro de la ecuación III a la actividad es $\sigma_p = -1,06$; ${}^2\Omega = 1,49$; ${}^1\epsilon = -1,02$ y $\pi = -0,05$. Es decir, el carácter electroceptor del grupo sustituyente en posición 3 sigue teniendo un peso importante en la actividad y la lipofilidad, aunque la posee, tiene una importancia pequeña en este modelo. Los dos índices de Estrada difieren en cuanto a su aporte y eso está dado por su naturaleza. El índice topológico ${}^1\epsilon$ se calcula a partir de la matriz de adyacencia de las aristas o enlaces del grafo químico y su orden 1 representa el conteo del número de enlaces sin tener en cuenta su multiplicidad; por lo tanto, debe lograrse al momento de la modelación, un balance entre la electronegatividad o carácter electroceptor del sustituyente a diseñar y su tamaño para equilibrar ambos efectos en el modelo.

Análisis de regresión de las muestras de N = 25, N = 34 y N = 36

Se decidió no hacer un aparte para cada una de las muestras pues como se verá por los resultados, después del último conjunto de derivados, los sustituyentes empleados se alejan del espacio de parámetros que sugieren los modelos.

Modelos encontrados para las muestras de N = 25, 34 y 36.

En el análisis de la muestra de N=25 (ec. V) aparece el término ${}^2\Omega_q$, correspondiente al índice topográfico basado en la matriz de adyacencia de los vértices ponderados por las densidades de carga, el cual es semejante a ${}^2\Omega$. Según el modelo, se mantienen las tendencias para σ_p y ${}^2\Omega$, en tanto ${}^2\Omega_q$ debe aumentar, esto se puede traducir en que se requiere un aumento en la polaridad del sustituyente. Sin embargo, ambos índices

topográficos se comportan en la muestra de diseño de manera contraria.

Para lograr una contribución positiva al aumento de la potencia (disminución del $\log MIC$), 2W debe disminuir y 2W_q debe aumentar, lo cual implicaría que simultáneamente disminuyeran los órdenes de enlace y aumentarían las densidades de carga sobre los átomos. En la muestra de 26 a 34 ($n = 10$), 2W aumenta en un 2,3 % mientras que 2W_q disminuye en aproximadamente 1.9 %. Esto significa que, desde un punto de vista estructural, la muestra resultado del diseño por el método PIH no se ajusta al modelo desarrollado por ARM y como tal, los resultados en cuanto a la respuesta biológica no fueron los esperados, a pesar del notable incremento promedio en el valor de s_p . Este resultado sugiere que, tal como se planteó en otra parte del trabajo, la reactividad de un compuesto no puede aumentar indefinidamente. Debe existir un balance entre los distintos aspectos a considerar como son carácter electrónico, tamaño, volumen, forma y lipofilidad. Grupos muy electroaceptores pueden convertir el compuesto en una entidad muy reactiva y por ende, inespecífica. Estos grupos pueden también incrementar su polaridad de forma tal que aumente su capacidad de asociación a macromoléculas polares disminuyendo su capacidad de paso a través de membranas con la consiguiente disminución de la actividad.

Al excluir dos *outliers*, se obtiene la ecuación VI en la que aparecen junto con σ_p dos índices topográficos que no estuvieron presentes en los modelos anteriores.

$$\log CMI = -2,893(\pm 0,429)\sigma_p + 0,700(\pm 0,120){}^2\Omega - 0,440(\pm 0,084){}^2\Omega_q + 1,852(\pm 0,204) \quad (V)$$

$$R = 0,900 \quad F_{(3,20)} = 28,483 \quad p < 0,000 \quad s = 0,435 \quad N = 24 \quad outlier = \text{Compuesto 15}$$

$$\log CMI = -3,050(\pm 0,347)\sigma_p + 1,087(\pm 0,223){}^4\Omega - 1,628(\pm 0,534){}^5\Omega_q + 2,220(\pm 0,172) \quad (VI)$$

$$R = 0,910 \quad F_{(3,19)} = 30,508 \quad p < 0,000 \quad s = 0,413 \quad N = 23 \quad outliers = \text{Compuestos 15 y 24}$$

$$\log CMI = -2,660(\pm 0,254)\sigma_p - 0,876(\pm 0,089){}^4\Omega + 2,164(\pm 0,123) + 1,745(\pm 0,407){}^5\Omega_{qc} + 0,128(\pm 0,059)\pi \quad (VII)$$

$$R = 0,915 \quad F_{(4,27)} = 34,611 \quad p < 0,000 \quad s = 0,372 \quad N = 32 \quad outliers = \text{Compuestos 15 y 24}$$

$$\log CMI = 2,460(\pm 0,281)\sigma_p + 0,555(\pm 0,114){}^2\Omega + 2,354(\pm 0,160) - 0,461(\pm 0,089){}^1\Omega_{qc} + 0,158(\pm 0,056)\pi \quad (VIII)$$

$$R = 0,906 \quad F_{(4,28)} = 32,090 \quad p < 0,000 \quad s = 0,383 \quad N = 33 \quad outliers = \text{Compuesto 24}$$

Ambos son de la serie Ω , uno de orden 4 y otro de orden 5, este último ponderado por las cargas. Esto señala un evidente cambio estructural en la muestra. Comienzan a tener una mayor importancia fragmentos de mayor tamaño. Sin embargo, la tendencia al aumento que se logra en el índice ${}^5\Omega_q$ no compensa el aumento del valor de ${}^4\Omega$. Los fragmentos de cinco átomos pesados con mayores densidades de carga —lo cual sugiere fragmentos más polares dentro de las moléculas— presentan en la muestra un menor peso en la actividad biológica que los fragmentos menos polares. Debe recordarse que la cubierta de las bacterias Gram (+) son más hidrófilas que las de las Gram (-), por lo que compuestos más polares deben tener mayor afinidad por dicha pared celular. Los índices topográficos ponderados por las densidades de carga, revelan por lo tanto, la mayor o menor importancia de este aspecto estructural sobre la actividad biológica dentro del conjunto de moléculas estudiadas.

La ecuación VII es “mejor” en algunos aspectos al compararla con la VIII. Ambas corresponden a la muestra de N = 34 y la diferencia fundamental estriba en la exclusión de un segundo *outlier* en la primera.

Esta exclusión permite aumentos discretos en los valores de R y F y una ligera disminución en el valor de σ . En esta muestra (N = 34), el tratamiento de los *outliers* hace que lo que resulta una aparente mejoría en el modelo desde el punto de vista de los estadígrafos de la ecuación, lleve implícitas varias dificultades en cuanto a la aceptación de uno u otro y lo que es más importante: la duda acerca de cuál, entre ambos, es el que

mejor refleja la muestra analizada. El compuesto 15 resultó *outlier* desde el primer momento del análisis de cada una de las muestras en la que está presente. El compuesto 24 se ha trabajado a partir del valor de actividad biológica estimado *a priori* por Boyd, dado que en la práctica no pudo ser sintetizado y solamente se empleó para disponer de ese punto en la recta de regresión. Debe suponerse que, en este punto del análisis, la muestra está alejándose del espacio de parámetros adecuado para lograr una respuesta biológica creciente.

A esta misma conclusión se puede llegar al examinar de la misma forma la ecuación IX, correspondiente a la muestra completa. En esta, también resultaron *outliers*, los compuestos 15 y 24.

Un aspecto interesante a notar es el hecho de que las variables ${}^p\Omega$ y ${}^p\epsilon$ en sus diferentes variantes, han sustituido a la refractividad molar en los modelos encontrados. Esto da una medida de que la variedad de ellas y su mejor capacidad descriptora de la estructura química les permite una mayor participación en las ecuaciones de regresión.

CONCLUSIONES

Los índices topográficos ${}^p\Omega$ y topológicos ϵ , constituyen descriptores de la estructura química adecuados para estudios QSAR. Ellos son capaces dentro de una muestra en estudio, de describir adecuadamente las modificaciones estructurales que en ella ocurran De igual manera, es posible mediante un análisis adecuado, predecir las transformaciones químicas que resulte necesario introducir en una molécula para lograr cambios específicos en el valor de los índices que conduzcan a

$$\log CMI = -2,078(\pm 0,244)\sigma_p + 0,738(\pm 0,151)^4\epsilon_p - 1,558(\pm 0,391)^5\Omega_{qc} + 2,078(\pm 0,127) \quad (IX)$$

$$R = 0,908 \quad F_{(3,30)} = 47,006 \quad p < 0,000 \quad s = 0,388 \quad N = 34 \quad outliers = \text{Compuestos 15 y 24}$$

una modificación orientada de la actividad biológica.

El análisis de regresión múltiple mantiene una vigencia innegable en los estudios QSAR, lo cual quedó demostrado en la comparación realizada con el método de proyección iterativa de hipersuperficies. Un manejo adecuado de los criterios de excusión de *outliers*, una selección adecuada de los descriptores y una correcta interpretación de los modelos hace que esta herramienta estadística sea de elección en este tipo de estudios.

BIBLIOGRAFIA

1. Kubinyi H., QSAR: Hansch Analysis and Related Approaches, Weinheim; New York; Basel; Cambridge; Tokyo; VCH. Eds. R. Mannhold, P. Krogsaard-Larsen y H. Timmerman, 1993.
2. Georgakopoulos C.G., Tsika O.G., Kiburis J.C. and Jurs P.C., *Anal. Chem.*, **63**, 18, 2025-2028, 1991.
3. Randic M., *J. Chem. Inf. Comp. Sci.*, **31**, 311-320, 1991.
4. Randic M., *New J. Chem.*, **15**, 7, 517-525, 1991.
5. Hansch C. and Fujita T., *J. Am. Chem. Soc.*, **86**, 1616-1626, 1964.
6. Kubinyi H. *J. Med. Chem.*, **19**, 587-600, 1976.
7. Cramer III R.D. and Bunce J.D., QSAR in Drug Design and Toxicology, (Proceedings del 6th European Symposium on QSAR, Portoroz, 1986, Hadzi, D., Jerman-Blazic, B. ed., Pharmacochem. Libr. 10, Elsevier, Amsterdam, 1987.
8. Randic M., *J. Am. Chem. Soc.*, **97**, 609, 1975.
9. Estrada E. Tesis Doctoral, Universidad Central de Las Villas, Cuba, 1995.
10. D.B.Boyd, *J. Med. Chem.*, **36**, 1443-49, 1993.
11. L.A.Casey, *et al.*, *J. Chem. Soc., Perkin Trans. 2*, 23-28, 1993.
12. a) Estrada E., *J. Chem. Inf. Comput. Sci.*, **35**, 31-33, 1995.
b) Estrada E., *J. Chem. Inf. Comput. Sci.*, **35**, 701-707, 1995.
c) Estrada E. and Montero L., *Mol. Eng.*, **2**, 363-373, 1993.
d) Estrada E., *J. Chem. Inf. Comput. Sci.*, **35**, 708-713, 1995.
13. Craig P.N., *J. Med. Chem.*, **14**, 680-684, 1971.
14. Cramer R.D., III, Snader K.M., Willis C.R., Chakrin L.W., Thomas J., Sutton B.M. *J. Med. Chem.*, **22**, 714-725, 1979.
15. HyperChem, Hypercube, Inc. and Autodesk, Inc. Release 3 for Windows, 1993 .
16. Dewar M.J.S., Zoebisch E.G., Healy E.F., Stewart J.J.P., *J. Am. Chem. Soc.*, **107**, 3902, 1985.
17. Stewart J.J.P., MOPAC 6.0. Manual QCPE # 581.
18. Stewart J.J.P. *J. Comp. Aid. Mol. Design*, **4**, 1, 1, 1990.
19. MOPAC V. 6 for 3/486/Pentium PC's. Release 1.02; March 1997; Windows 95 and NT environments. Universidad de la Habana, Cuba, 1994-97
20. a) Brodyden C. G. *J. Inst. Math. Appl.*, **6**, 222, 1970.
b) Fletcher R. *Comp. J.*, **13**, 317, 1970. Goldfarb D. *Math. Comp.*, **24**, 23, 1970.
c) Shanno D.F. *Math. Comp.*, **24**, 647, 1970.
21. StatSoft, Inc. STATISTICA for Windows [Computer program manual]. Tulsa, OK: StatSoft, Inc., 2325 East 13th Street, Tulsa, OK 74104, (918) 583-4149, fax: (918) 583-4376, 1995.
22. Balbes L.M., Wayne Mascarela S. and Boyd D.B., *Reviews in Computational Chemistry*, Vol 5, 337, 1994.



ACTIVIDADES CIENTIFICAS
MINISTERIO DE EDUCACION SUPERIOR DE CUBA

III TALLER INTERNACIONAL SOBRE AGRICULTURA SOSTENIBLE EN CONDICIONES DE MONTAÑA

Centro Universitario de Guantánamo.

Del 22 al 25 de abril del 2004.

Temáticas:

Desarrollo de la Montaña, Cultivos Múltiples, Desarrollo Rural Sostenible, Sistema de Producción Agroalimentario y La Enseñanza en la Montaña.

Para mayor información:

Dr. José Machuca, Comité Organizador. E-mail: drigtm@cug.co.cu