

Aplicación en QSPR del parámetro topográfico V_{c3} a sustancias derivadas de la benceno sulfonamida

Recibido el 9 de febrero de 2007

E. CORNWELL*

*Departamento de Química Inorgánica y Analítica.
Facultad de Ciencias Químicas y Farmacéuticas.
Universidad de Chile*

RESUMEN

Un nuevo parámetro topográfico (V_{c3}) aplicado a la disciplina QSPR se utiliza para la reducción del número de variables independientes a una sola variable denominada V_{c3} . La reducción está basada en la relación de las distancias Euclidianas. Este procedimiento fue aplicado a un modelo de 19 sustancias derivadas del benceno sulfonamida utilizando el benceno sulfonamida como sustancia referente. Estas sustancias fueron caracterizadas por un grupo de propiedades fisicoquímicas en la forma de variables independientes, las variables utilizadas fueron: el índice de refracción, la tensión superficial y una variable auxiliar indicadora de la presencia o ausencia de átomos de cloro en la molécula bajo estudio, se utilizó como variable dependiente la constante de disociación pK_a obteniéndose un modelo simple de regresión lineal correspondiente a: $pK_a = m \cdot V_{c3} + n$, expresión más coherente que su contraparte correspondiente a la regresión multivariable. La aplicación del uso de la reducción de variables elimina el problema del proceso de ortogonalización de las variables o el uso de la Componentes Principales (PCA) para obtener modelos consistentes.

Palabras clave: Reducción.—Variables independientes.—Distancias euclidianas.

Información de contacto:

Departamento de Química Inorgánica y Analítica. Facultad de Ciencias Químicas y Farmacéuticas. Universidad de Chile, Olivos 1007, Santiago, Chile. Fax: 56-2-7370567. Teléfono: 9782851.
e-mail: ecornwel@abello.dic.uchile.cl

ABSTRACT

Application of V_{c3} topographic QSPR parameters to benzene sulfonamida substances derivatives

A novel QSPR topographic parameter (V_{c3}) is used for independent variables reduction number to one. V_{c3} consisted in Euclidian distances relations. This procedure was applied to a model of 19 benzene sulfonamide derivatives substances using benzene sulfonamide like referent substance characterized by physico-chemicals properties used like independent variables, this variables are: refraction index, surface tension and an extra dummyvariable that indicated the presence or absence chlorine atoms in the study molecule, pKa was used like dependent variable. The linear regression proposed is of the form. $pKa = m \cdot V_{c3} + n$ that is more coherent than its counterpart multivariable regression. The used of this variable reduction eliminated problems of orthogonal procedure or the used Principal Component Analysis (PCA) to obtain consistent models.

Key words: Reduction.—Independent variables.—Euclidian distances.

INTRODUCCIÓN

Un novedoso parámetro V_{c3} utilizado como modelo matemático en la disciplina QSPR (Relación propiedad estructura química) es propuesto por el autor como un mecanismo general para reducir el número de variables independientes en las regresiones multivariadas.

El propósito de este artículo es extender esta idea propuesta para su aplicación en estructuras más complejas que los hidrocarburos saturados tratados con antelación (1) para obtener modelos más simples definidos como relaciones lineales con una variable V_{c3} independiente. Para este caso corresponde a la forma general $y = m \cdot V_{c3} + n$ cuya variable independiente definida como V_{c3} para este caso expuesto es función de propiedades fisicoquímicas propias de las sustancias a tratar. El parámetro V_{c3} fue exitosamente aplicado a un grupo de 35 hidrocarburos saturados (1). En el presente trabajo se utilizó un número de 18 derivados del benceno sulfanilamida utilizando el benceno sulfonamida como sustancia de comparación. Cada derivado está representado por un conjunto ordenado $[X_i, Y_i, Z_i]$ en el cual cada elemento en mayúscula representa una propiedad fisicoquímica y el índice i representa una sustancia en particular. Para la sustancia de comparación (benceno sulfonamida) se define el conjunto ordenado

$[X_0, Y_0, Z_0]$. En ambos conjuntos, X representa el índice de refracción (IR), Y representa la tensión superficial (ST) y Z representa la variable auxiliar (2) (I) cuyo valor adquiere un valor 1 o 0 de acuerdo si la molécula contiene en su estructura átomos de cloro o no los contiene, respectivamente. Se estableció que el uso de otro conjunto distinto al establecido como $[X_0, Y_0, Z_0]$ en base a el benceno sulfonamida, como ejemplo, usar otra sustancia entre los derivados de esta sustancia señalada, conllevan a la obtención de regresiones de menor calidad estadísticas que la obtenida mediante el conjunto señalado. La causa de ello, es probablemente debida a las diferencias significativas entre esta sustancia base (benceno sulfonamida) y sus derivados en cuanto a las estructuras moleculares y las distribuciones electrónicas que las caracterizan.

El modelo propuesto en este trabajo se define por la relación $pK_a = m^* V_{c3} + n$, siendo la variable independiente la distancia Euclidiana entre los pares ordenados de los conjuntos $[X_i, Y_i, Z_i]$ y $[X_0, Y_0, Z_0]$.

Se debe indicar que el modelo de regresión multivariable que relaciona pK_a con las propiedades fisicoquímicas de esta sustancias bajo estudio presentan como en todo caso general de este tipo de regresiones multivariables magnitudes y signos de los parámetros multiplicativos de cada término contributivo del modelo incorrectos, que no interpretan el comportamiento de cada término en el total de la ecuación aún cuando el modelo sea adecuado para la obtención de las variables dependientes calculadas. Para tener concordancia total de estos modelos se hace necesario un proceso de ortogonalización de sus variables independientes (3). Además de ello, el número de variables independientes utilizados en las regresiones multivariables deberá estar de acuerdo con el número de sustancias presentes que conforman el modelo. En caso contrario se obtienen coeficientes de determinación estadísticos (R^2) falsos por exceso (4) que no tienen proporción con las diferencias entre los valores experimentales y calculados por el modelo. Estas limitaciones encontradas en las regresiones multivariables no existen para el modelo propuesto, los parámetro m , n quedan bien definidos como elementos de la recta representativa de la regresión lineal y el signo de la derivada de su función es consecuente con los signos de las derivadas correspondientes a cada elemento contenido en V_{c3} .

Los procesos químicos, fisiológicos, y patológicos en los cuales la constante de disociación pK_a está involucrada están ampliamente investigados para el grupo de los derivados de las benceno sulfonamidas (2). Además, se ha demostrado que algunas derivados de las benceno sulfonamidas son agentes químicos importantes y se usan estas sustancias para el tratamiento de úlceras gastro-intestinal-duodenal, desórdenes neurológicos, glaucoma, debilidad causada por la altura, y algunas formas de tumor (2).

PROCEDIMIENTO

Las sustancias bajo estudio se caracterizaron mediante un grupo de seis propiedades fisicoquímicas. Éstas son: el índice de refracción molar (MRI), el volumen molar, la polarizabilidad, el índice de refracción molar (IR), la tensión superficial (ST) y la variable auxiliar. Todos estos datos se obtuvieron de literatura (2) (ver Tabla 1). De este conjunto de propiedades se eligió un grupo de tres propiedades, inherente a las sustancias con el cual se logró un modelo de regresión multivariable óptimo evaluado y comparado mediante los parámetros estadísticos comunes. Estos son el índice estadístico de determinación (R^2), la desviación estándar de la regresión (s. d.) y el índice de Fisher (F). El número de variables independientes ocupadas está en proporción con el número de casos (19 sustancias), de acuerdo a la bibliografía ya citada (4). La regresión óptima se eligió de un conjunto de 20 regresiones que implica el total de combinaciones posibles calculadas como: $C_3^6 = 6! / 3! 3! = 20$ regresiones con distintas combinaciones de a tres propiedades fisicoquímicas. La tríada óptima de la regresión multivariable óptima obtenida fue: [IR, ST, I]. La estructura particular de la regresión multivariable se señala en la ecuación 1:

$$pK_a = -21.2724 (\pm 25.2200) + 24.3608 (\pm 16.7276)*IR - 0.1604 (\pm 0.0444)*ST - 0.5955 (\pm 0.2476)*I \quad (1)$$

$$R^2 = 89.1946$$

$$s. d. = 0.3351$$

$$F = 44.0200$$

El parámetro V_{c3} se obtuvo para cada derivado de la benceno sulfonamida a través de la relación de distancia Euclidiana entre el conjunto ordenado $[X_i, Y_i, Z_i]$ y el de referencia $[X_0, Y_0, Z_0]$ propio de la benceno sulfonamida, cuya estructura particular es [1.645, 60.700, 0]. La distancia para cada sustancia se calcula mediante la ecuación 2:

$$V_{c3} = [(X_i - X_0)^2 + (Y_i - Y_0)^2 + (Z_i - Z_0)^2]^{1/2} \quad (2)$$

Esta ecuación fue aplicada a los 19 derivados y los valores de V_{c3} se señalan en la Tabla 1. La regresión entre pKa y V_{c3} corresponde a la ecuación 3:

$$pKa = 9.0556 (\pm 0.1252) - 0.1061 (\pm 0.0104) * V_{c3} \quad (3)$$

$$r = -0.9235$$

$$s. d. = 0.3695$$

$$F = 104.39$$

Para el caso de la relación multivariable (1), la regresión pKa experimental vs. pKa calculado mediante la regresión 1 proporciona los siguientes parámetros estadísticos ($r = 0.9448$, s. d. = 0.3154, $F = 149.94$).

Para el caso de la relación (3) la regresión pKa experimental vs. pKa calculado mediante la regresión 3 proporciona los siguientes parámetros estadísticos ($r = 0.9237$, s. d. = 0.369, $F = 104.51$).

En la Tabla 1 se señalan tanto valores de pKa calculados a través de la regresión multivariable y sus errores respecto a los experimentales como los valores de pKa calculados mediante la relación 3 y sus errores respecto a los experimentales. Los pKa calculados para ambos casos, regresión multivariable y en función de V_{c3} no presentan diferencias estadísticamente para un 95% de confianza entre sus promedios, desviación estándares, mediana (prueba de: Mann-Witney W.) y distribución (prueba de Kolgomorov-Smirnov). Esto implica que no existen diferencias sustantivas entre ambos métodos de regresión, en cuanto a la evaluación de los pKa calculados.

Las regresiones y evaluaciones estadísticas se llevaron a término mediante los software Statgraphic y Origin (5, 6).

DISCUSIÓN Y CONCLUSIONES

El método propuesto de regresión mediante la reducción de variables V_{c3} es más coherente con las funciones de regresión para el modelo lineal $y = m \cdot x$ de pKa versus el índice de refracción, pKa versus la tensión superficial y pKa versus variable independiente auxiliar (I) que las regresiones multivariantes. Se tiene que: la derivada $d(pKa)/d(IR)$ para la función pKa vs. (IR) es igual a -40.6151 , de la misma forma $d(pKa)/d(ST)$ para la función pKa vs. (ST) es igual a -0.1061 y $d(pKa)/d(I)$ para la función pKa vs. (I) es igual a -0.536 . Siendo la derivada $d(pKa)/d(V_{c3}) = -0.1061$ se puede deducir que existe coherencia en los signos con las ecuaciones por separado, más aún, el modelo propuesto V_{c3} indica un mismo valor de derivada que $d(pKa)/d(ST)$ para la función pKa vs. (ST) señalando que la variable independiente (ST) sería la variable de mayor peso para un modelo multivariable. Si se evalúan las derivadas parciales del modelo multivariable se tiene que $(\delta(pKa)/\delta ST)_{IR,1} = -0.1604$ y $(\delta(pKa)/\delta ST)_{IR,ST} = -0.5955$, manteniéndose cierta similitud con los modelos con una sola variable independiente pero, $d(pKa)/d(IR)_{ST,1} = 24.3608$ valor que es incoherente en magnitud y signo con $d(pKa)/d(IR) = -40.6151$, el segundo término de la regresión multivariable no explica ni el signo verdadero ni la magnitud de la contribución de la variable (IR) y en aquella relación, se compensa este factor positivo (24.3608) por medio de la magnitud y signo del término independiente (-21.2724) de tal forma que los valores calculados de pKa mediante un proceso de regresión multivariable y aquellos calculados por el método propuesto V_{c3} son estadísticamente equivalentes, como se indicó anteriormente. Este hecho particular señala la necesidad de un proceso de ortogonalización, reducción de variables no significativas por el método PCA o como alternativa la reducción de variables por el método propuesto. Se entiende en todo el texto como reducción no una eliminación si no una función de funciones.

Es necesario señalar que este procedimiento estudiado amerita mayores análisis pero desde ya se puede decir que es un método más sencillo que los métodos de ortogonalización o PCA.

TABLA 1. Sustituyentes químicos y parámetros fisicoquímicos usados en este estudio

Núm.	Sustituyente	pKa Exp.	MRI cm ³	Volumen molar	Polariza- 10 ⁻⁴ cm ²	I.R.	ST dones/cm	J	V _s	pKa Calculada Ec. Multivar.	error %	pKa Calculada Ec. Función V _s	error %
1	3,5-di-NO ₂	6,19	79,61	204	31,56	1,706	88,5	0	27,8001	6,09	1,61	6,11	1,37
2	3-NO ₂ -5-Cl	6,92	78,4	204,6	31,08	1,692	77,6	1	16,9296	6,90	0,25	7,26	-4,90
3	4-NO ₂	6,97	73,58	192,6	29,17	1,689	76,8	0	16,1001	7,55	-8,37	7,35	-5,41
4	3-NO ₂ -4-Cl	7,16	78,4	204,6	31,08	1,692	77,6	1	16,9296	6,90	3,60	7,26	-1,38
5	4-CN	7,36	72,11	190,7	28,58	1,680	74,7	0	14,0001	7,67	-4,22	7,57	-2,85
6	4-SO ₂ -NH ₂	7,45	79,98	209,7	31,7	1,687	76,7	0	16,0001 ^I	7,52	-0,95	7,36	1,24
7	3,5-di-Cl	7,54	77,2	204,7	30,6	1,678	67,8	1	7,1702	8,13	-7,87	8,29	-10,01
8	3-NO ₂	7,67	73,58	192,6	29,17	1,689	76,8	0	16,1001	7,55	1,52	7,35	4,21
9	3-SO ₂ -NH ₂	7,81	79,98	209,7	31,7	1,687	76,7	0	16,0001	7,52	3,71	7,36	5,79
10	3-CN	7,83	72,11	190,7	28,58	1,680	74,7	0	14,0001	7,67	2,03	7,57	3,32
11	3-Cl	8,28	72,37	192,7	28,69	1,674	66,4	1	5,7871	8,26	0,23	8,44	-1,95
12	3-CO-Me	8,34	76,91	212,3	30,49	1,644	62,3	0	1,6000	8,78	-5,31	8,89	-6,54
13	4-Cl	8,56	72,37	192,7	28,69	1,674	66,4	1	5,7871	8,26	3,50	8,44	1,39
14	3-O-Me	8,72	73,91	204,8	29,3	1,641	60,7	0	0,0000	8,97	-2,83	9,06	-3,85
15	H	8,97	67,54	180,8	26,77	1,641	60,7	0	0,0000	8,97	0,04	9,06	-0,95
16	3-N-Me ₂	9,01	80,68	218,8	31,98	1,658	63,6	0	2,9000	8,92	1,05	8,75	2,91
17	3-Me	9,05	71,89	205,8	28,5	1,653	61,3	0	0,6001	9,16	-1,25	8,99	0,64
18	4-Me	9,25	72,17	197,1	28,61	1,653	61,3	0	0,6001	9,16	0,94	8,99	2,79
19	4-O-Me	9,34	73,91	204,8	29,3	1,641	60,7	0	0,0000	8,97	4,00	9,06	3,05
20	4-N-Me ₂	9,46	80,68	218,8	31,98	1,658	63,6	0	2,9000	8,92	5,76	8,75	7,53

I, indica una variable independiente auxiliar. Se define este parámetro con un valor 1 para moléculas cloradas, en caso contrario su valor es 0. MRI implica el índice de refracción molar.

BIBLIOGRAFÍA

- (1) CORNWELL, E. (2006): New idea for the topological index evaluation and treatise multiple regression with three independent variables. Saturated hydrocarbons used like a model. *J. Chil. Chem. Soc.* 50: 765-768.
- (2) THAKUR, A. (2005): QSAR study on benzenesulfonamide dissociation constant pKa: physicochemical approach using surface tension. *Arkivoc* XIV 49-58.
- (3) RANDIC, M. (1997): On characterization of chemical structure. *J. Chem. Inf. Comput. Sci.* 37: 672-687.
- (4) TOPLIS, J. C. and EDWARDS, R. P. (1979): Chance factors in studies of quantitative structure activity relationships. *J. Med. Chem.* 22: 1238-1244.
- (5) Statgraphic Plus 5.1 Copyrighty 1994-2001 Statistically Graph Corp.
- (6) Origin 7312117. 0301 (B30019) Copyright 1991-2002 Origin Lab. Corporation. One Round Plaza Northampton, MA 01060 USA.