

ESTUDO DO COEFICIENTE DE PARTIÇÃO OCTANOL-ÁGUA DE BIFENILAS POLICLORADAS (PCBs) UTILIZANDO PARÂMETROS TOPOLÓGICOS

Lucicleide Ribeiro da Silva

Departamento de Química, Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora, 36036-330 Juiz de Fora - MG

Márcia M. C. Ferreira*

Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970 Campinas - SP

Recebido em 19/3/02; aceito em 9/10/02

A STUDY OF OCTANOL/WATER PARTITION COEFFICIENT OF POLYCHLORINATED BIPHENYLS (PCBs) USING TOPOLOGICAL INDICES. A model based on chemical structure was developed for the accurate prediction of octanol/water partition coefficient (K_{ow}) of polychlorinated biphenyls (PCBs), which are molecules of environmental interest. Partial least squares (PLS) was used to build the regression model. Topological indices were used as molecular descriptors. Variable selection was performed by Hierarchical Cluster Analysis (HCA). In the modeling process, the experimental K_{ow} measured for 30 PCBs by *thin-layer chromatography - retention time* (TLC-RT) has been used. The developed model ($Q^2 = 0,990$ and $r^2 = 0,994$) was used to estimate the $\log K_{ow}$ values for the 179 PCB congeners whose K_{ow} data have not yet been measured by TLC-RT method. The results showed that topological indices can be very useful to predict the K_{ow} .

Keywords: topological indices; chemometrics; QSPR.

INTRODUÇÃO

Os PCBs possuem fórmula geral $C_{12}H_nCl_y$, com $n=0,\dots,10$ e $y=0,\dots,10$. Desta maneira, existem 209 congêneros de PCBs organizados em subcategorias, os homólogos. Estes, por sua vez, são formados por diferentes congêneros que tenham o mesmo número de átomos de cloro¹. O número de identificação de cada congênero foi criado por Ballschmiter e Zell², e mais tarde adotado pela IUPAC. Outra maneira de se referir a cada composto é indicando a posição do átomo de cloro no anel, de acordo com a Figura 1, seguida pelo termo CB (ex. 2,2',3-CB).

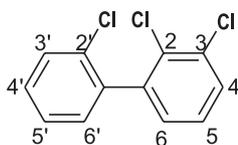


Figura 1. Esquema da molécula 2,2',3-CB

Os PCBs são largamente utilizados na indústria. A versatilidade industrial deles está diretamente relacionada às suas propriedades físicas, tais como resistência à oxidação e redução, excelente propriedade como isolante elétrico, estabilidade térmica e não inflamabilidade³. Apesar dessas propriedades físicas importantes, juntamente com seu amplo uso, o descarte dos PCBs é inapropriado, resultando na contaminação do ecossistema global. Nos seres humanos os problemas de saúde que podem ocorrer dependem da dose de exposição⁴. A natureza lipofílica e a persistência dos PCBs contribuem para seu alto potencial de bioacumulação nos níveis mais altos da cadeia alimentar. Isto é, resíduos de PCBs são rotineiramente detectados em peixes e tecido adiposo humano⁵.

Aceita-se cada vez mais que o único método rigoroso para resolver os problemas causados pelos PCBs no meio ambiente consiste em obter as propriedades de todos os congêneros, e tratar cada

congênero separadamente. Desta forma, estudar a dinâmica dos PCBs no meio ambiente exige a medida experimental de vários parâmetros. Infelizmente, poucos congêneros são sintetizados em quantidade suficiente para que seja possível a determinação de suas propriedades⁶, em função das dificuldades experimentais e do alto custo envolvido. Desta forma, há grandes dificuldades em se obter muitas das propriedades de vários PCBs, e a literatura é escassa em relação a estes dados. Daí a importância e necessidade de busca por modelos teóricos para previsão das mesmas.

O coeficiente de partição octanol-água (K_{ow}) dos PCBs foi a propriedade selecionada para este estudo. K_{ow} é uma medida da lipofilicidade de um composto e é definido como a razão da concentração do mesmo, no equilíbrio, após dissolução em um sistema de duas fases, formadas por dois solventes imiscíveis, água e octanol. A determinação de K_{ow} tem sido recomendada pelo programa de evolução de compostos perigosos (OECD)⁷. Nos últimos anos, K_{ow} vem sendo muito utilizado em várias áreas, com um número elevado de publicações divulgando a sua correlação com outras propriedades físicas, químicas e biológicas dos compostos, além do sucesso de seu uso na estimativa de fatores de bioconcentração, coeficiente de partição carbono-orgânico, toxicidade e solubilidade⁸. Isto se deve ao fato de K_{ow} estar relacionado com a interação do composto em estudo com o meio, no que diz respeito à absorção e transporte. Entre as técnicas experimentais utilizadas para medir K_{ow} se destacam: (i) "shake-flask", que é o método clássico de determinação de K_{ow} ; (ii) técnica geradora de coluna (GC); (iii) medidas de cromatografia líquida de fase reversa de alta eficiência e de camada fina (RP-HPLC e RP-TLC). Todas estas técnicas apresentam suas limitações, como a absorção e volatilização do soluto (GC)⁹, ou a dificuldade analítica na determinação de K_{ow} de substâncias pouco solúveis, pois a determinação da concentração de PCB em água pode ser imprecisa devido à formação de agregados moleculares¹⁰. Neste trabalho, optou-se por utilizar valores experimentais de K_{ow} extraídos da literatura¹¹ que fossem medidos pela mesma técnica experimental. Foram encontrados valores experimentais de K_{ow} para 30 PCBs obtidos pela medida do tempo de retenção por cromatografia em camada fina (RT-TLC)¹¹.

*e-mail: marcia@iqm.unicamp.br

K_{ow} pode também ser calculado teoricamente utilizando diferentes metodologias: (i) métodos baseados em fragmentos da molécula. Neste caso, $\log K_{ow}$ é estimado a partir da contribuição aditiva de grupos funcionais e fragmentos anexados à base molecular¹²; (ii) métodos baseados na contribuição atômica¹³, em que cada átomo tem uma contribuição aditiva para o $\log K_{ow}$ da molécula. No caso dos PCBs, estes métodos não produzem resultados precisos, uma vez que os valores obtidos de $\log K_{ow}$ para PCBs de um mesmo grupo homólogo têm o mesmo valor; (iii) métodos baseados em modelos que utilizam parâmetros obtidos por meio de cálculos quânticos¹⁴. Estes métodos apesar de produzirem bons resultados, necessitam de programas e cálculos sofisticados para a otimização da geometria molecular e obtenção dos parâmetros estruturais, eletrônicos, energéticos etc., a serem quantitativamente correlacionados com K_{ow} . Com base no exposto acima, é visível a importância de se utilizar uma metodologia alternativa para o cálculo de K_{ow} que seja eficiente e simples. Neste trabalho, foram utilizados alguns parâmetros topológicos das moléculas de PCBs como variáveis, pois como será mostrado adiante, estes parâmetros são de fácil obtenção, não exigem programas específicos e sofisticados, nem o mesmo tempo computacional que os parâmetros obtidos de cálculos quânticos.

O estudo das relações quantitativas entre a estrutura química de uma molécula ou uma série de moléculas e sua atividade biológica (QSAR), ou de relações entre a sua estrutura e alguma propriedade físico-química (QSPR), é de grande importância na bioquímica, química medicinal e ambiental. Estes estudos são importantes pois podem fornecer informações sobre as propriedades e características de compostos já existentes, além de subsídios para propor novos compostos com as características de interesse. Esta área teve seu grande avanço na década de 1960, com o trabalho pioneiro de Hansch¹⁵ e desde então tem sido intensamente aprimorada e utilizada.

Os modelos de estrutura-atividade/propriedade fundamentam-se na hipótese de que a estrutura molecular contém as características responsáveis pelas propriedades físicas, químicas e biológicas do composto em questão. Com o avanço tecnológico no que diz respeito a novos computadores e programas, a obtenção destes modelos tornou-se uma alternativa bastante atraente e promissora¹⁶.

A metodologia geral se inicia com o cálculo de uma série de parâmetros (descritores), onde a grande maioria é obtida a partir da estrutura molecular. O passo posterior ao cálculo dos descritores é a construção de um modelo capaz de relacioná-los com a propriedade/atividade de interesse. A ferramenta escolhida neste caso é a quimiometria, que dispõe de métodos para a análise de dados multivariados. Os métodos quimiométricos utilizados foram HCA (Hierarchical Cluster Analysis) e PLS (Partial Least Squares)¹⁷.

Neste trabalho, foi estudada a relação entre estrutura e propriedade de PCBs, aplicando métodos quimiométricos e utilizando como descritores alguns índices topológicos. Os descritores topológicos são obtidos a partir da topologia da molécula e formam um importante conjunto, freqüentemente utilizado em estudos de toxicidade e propriedades biomédicas¹⁸. Estes parâmetros químicos foram desenvolvidos utilizando a teoria de grafos¹⁹. Em termos gerais, eles caracterizam a molécula por meio de um único número, e quantificam a estrutura molecular descrevendo características como disposição dos átomos na molécula, forma e ramificação da mesma.

PARTE EXPERIMENTAL

Parâmetros topológicos

Para a representação de grafos das moléculas, os átomos são representados como vértices e as ligações como traços²⁰. A Figura 2 mostra a representação de grafo do 2-pentanol que contém 5 átomos de

carbono, 1 átomo de oxigênio e 6 vértices. A seguir são descritos os índices topológicos que foram utilizados neste trabalho.

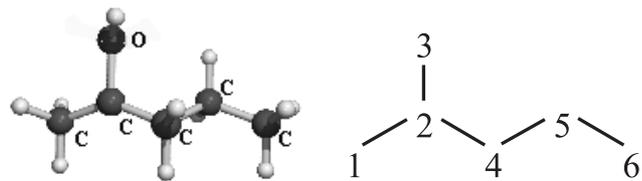


Figura 2. Molécula de 2-pentanol e sua representação de grafo

Índice de Wiener

O índice de Wiener foi introduzido em 1947²¹, sendo o primeiro índice de grafo introduzido na química. Para definir o índice de Wiener, consideremos uma representação de grafo de uma molécula constituída de k átomos onde v_1, v_2, \dots, v_k são os respectivos vértices. A distância entre dois átomos i e j , D_{ij} , é dada pela distância entre os vértices v_i e v_j que é igual ao número de traços (ligações) considerando-se o menor caminho que conecte v_i e v_j . Por exemplo, a menor conexão entre os átomos 1 e 3 da Figura 2 corresponde a dois traços. Portanto $D_{13} = 2$. Calculando-se as distâncias entre todos os átomos constituintes da molécula, constrói-se a matriz de distâncias \mathbf{D} (6x6) que é dada por:

$$\mathbf{D} = \begin{bmatrix} 0 & 1 & 2 & 2 & 3 & 4 \\ 1 & 0 & 1 & 1 & 2 & 3 \\ 2 & 1 & 0 & 2 & 3 & 4 \\ 2 & 1 & 2 & 0 & 1 & 2 \\ 3 & 2 & 3 & 1 & 0 & 1 \\ 4 & 3 & 4 & 2 & 1 & 0 \end{bmatrix} \quad (1)$$

À direita da matriz tem-se o resultado da soma de cada linha, formando uma nova coluna. A soma desta coluna resultante, dividida por dois, é igual ao índice de Wiener, definido na Equação 2 como:

$$W = \frac{1}{2} \sum_{ij} (D_{ij}) \quad (2)$$

Para a molécula da Figura 2, o resultado obtido para o índice de Wiener foi $W = 32$.

Índice de conectividade

O conceito de índice de conectividade molecular foi introduzido por Randić²². A idéia é quantificar a estrutura molecular de maneira que fragmentos e subestruturas sejam incorporados como índices numéricos. Primeiro calcula-se, para cada átomo i , o descritor de valência (δ^v) e o descritor de elétrons sigma δ , ambos definidos nas Equações 3 e 4 respectivamente.

$$\delta^v = \frac{Z^v - h}{Z - Z^v - 1} \quad (3)$$

$$\delta = \sigma - h \quad (4)$$

σ é o número de elétrons sigma, Z^v é o número de elétrons de valência, Z é o número atômico e h o número de átomos de H ligados ao átomo i .

Na molécula da Figura 2, de acordo com as Equações 3 e 4, os valores de δ e δ^v são:

$$\begin{aligned} \delta_{(\text{CH}_3)} &= \delta_1 = \delta_6 = 4 - 3 = 1; \delta_{3(\text{-OH})} = 2 - 1 = 1; \\ \delta_{(\text{CH}_2)} &= \delta_2 = 4 - 1 = 3; \delta_{(\text{CH}_2)} = \delta_4 = \delta_5 = 4 - 2 = 2; \\ \delta_1^v &= \delta_6^v = (4-3)/(6-4-1) = 1; \delta_3^v = (6-1)/(8-6-1) = 5; \\ \delta_2^v &= (4-1)/(6-4-1) = 3; \delta_4^v = \delta_5^v = (4-2)/(6-4-1) = 2. \end{aligned}$$

Os índices de conectividade molecular são simbolizados por ${}^m\chi_t$. As subestruturas de uma molécula e a ordem do índice de conectividade são definidas pela fragmentação da representação de grafo:

- a - átomos (ordem zero, $m = 0$);
- b - uma ligação (primeira ordem, $m = 1$);
- c - fragmentos de duas ligações consecutivas (segunda ordem, $m = 2$);
- d - três ligações contínuas (terceira ordem de caminho, $m = 3, t = P$);
- e - fragmentos incluindo ramificações, três átomos ligados a um átomo central ($m = 3, t = C$);
- f - fragmentos incluindo caminho/ramificação, referem-se a quatro ligações adjacentes, com 3 ligações em um átomo central, equivalente à estrutura do isopentano ($m = 4-6, t = PC$);
- g - fragmento de cadeia, com anéis de 3, 4, 5, ...átomos ($m = 3-6, \dots, t = CH$).

O índice de conectividade de valência ${}^m\chi_t$ é calculado multiplicando-se os valores de δ de cada átomo da subestrutura. O inverso da raiz quadrada deste produto é somado com todos os outros produtos de subestruturas de ordem m e tipo t :

$${}^m\chi_t = \sum (\prod_{k=1}^{m+1} \delta_k)^{-0.5} \quad (5)$$

O índice de conectividade de valência é obtido usando-se o valor de δ^v no lugar de δ .

Vejam agora alguns exemplos do cálculo de ${}^m\chi_t$ para a molécula da Figura 2. Para o cálculo de ${}^1\chi$ são considerados os átomos adjacentes existentes na molécula. Os pares são: 1-2, 2-3, 2-4, 4-5 e 5-6. Os valores de δ utilizados já foram calculados acima. Assim:

$${}^1\chi = (\delta_1\delta_2)^{-0.5} + (\delta_2\delta_3)^{-0.5} + (\delta_2\delta_4)^{-0.5} + (\delta_4\delta_5)^{-0.5} + (\delta_5\delta_6)^{-0.5} = 2,770$$

Quando se calcula o índice de ordem 2 (${}^2\chi$), deve-se considerar duas ligações consecutivas e, portanto, 3 átomos. Para a molécula em questão, as combinações são: 1-2-3, 1-2-4, 3-2-4, 2-4-5, 4-5-6. Os produtos de δ dos três átomos são somados, resultando em:

$${}^2\chi = (\delta_1\delta_2\delta_3)^{-0.5} + (\delta_1\delta_2\delta_4)^{-0.5} + (\delta_2\delta_3\delta_4)^{-0.5} + (\delta_2\delta_4\delta_5)^{-0.5} + (\delta_4\delta_5\delta_6)^{-0.5} = 2,183$$

No cálculo do índice de conectividade de ramificação é necessário identificar no grafo subestruturas semelhantes ao isobutano, em que três átomos estão ligados a um átomo central (Y). Neste exemplo existe apenas uma subestrutura, e esta é formada pelos átomos 1-2-3-4:

$${}^3\chi_c = (\delta_1\delta_2\delta_3\delta_4)^{-0.5} = 0,408$$

Para se obter o índice de caminho/ramificação de ordem 4, procura-se na molécula subestruturas que sejam semelhantes à do isopentano (C). No caso da molécula deste exemplo, tem-se apenas um fragmento com esta forma, constituído pelos átomos 1-2-3-4-5. Deve-se notar que este índice procura estruturas semelhantes ao isopentano, mesmo que envolva outros átomos, com exceção do átomo de hidrogênio. Para a molécula considerada, o índice de ordem quatro é:

$${}^4\chi_{pc} = (\delta_1\delta_2\delta_3\delta_4\delta_5)^{-0.5} = 0,289$$

Índice de Harary

O índice de Harary (H_D) é obtido a partir da matriz de distância \mathbf{D} , e recebeu este nome em homenagem a Frank Harary²³. H_D é definido como:

$$H_D = \frac{1}{2} \sum_{ij} \frac{1}{D_{ij}} \quad (6)$$

Os valores de D_{ij} são os elementos da matriz \mathbf{D} obtidos quando se calculou o índice de Wiener (Equação 1). Então,

$$\frac{1}{\mathbf{D}} = \begin{bmatrix} 0 & 1 & 1/2 & 1/2 & 1/3 & 1/4 \\ 1 & 0 & 1 & 1 & 1/2 & 1/3 \\ 1/2 & 1 & 0 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1 & 1/2 & 0 & 1 & 1/2 \\ 1/3 & 1/2 & 1/3 & 1 & 0 & 1 \\ 1/4 & 1/3 & 1/4 & 1/2 & 1 & 0 \end{bmatrix} \begin{matrix} 2,583 \\ 3,833 \\ 2,583 \\ 3,500 \\ 3,167 \\ 2,333 \end{matrix}$$

Os valores à direita da matriz referem-se à soma de cada linha. A soma deles dividida por dois, é igual ao índice de Harary e, neste caso, $H_D = 9$. Este índice é caracterizado por proporcionar uma contribuição menor das ligações terminais do que das ligações internas, ao contrário do índice de Wiener.

Índice eletrotológico

Os índices eletrotológicos são valores numéricos calculados para cada átomo da molécula que contém informações sobre o ambiente topológico dos átomos e sobre a interação eletrônica devido a outros átomos da molécula²⁴.

Para se obter o valor do estado eletrotológico (ES_i) de um átomo, é necessário primeiro se obter o valor do estado intrínseco I_i

$$I_i = \frac{\left[\left(\frac{2}{n} \right)^2 \delta_i^v + 1 \right]}{\delta_i} \quad (7)$$

onde δ_i^v e δ_i foram definidos nas Equações 3 e 4 respectivamente, e n é o número quântico principal.

A perturbação no átomo i , originária da presença de todos os outros átomos j da molécula, é uma função da diferença entre os valores intrínsecos e da distância entre i e j , r_{ij} , definida como o número de átomos entre i e j (incluindo i e j). Esta perturbação é definida na Equação 8

$$\Delta I_i = \sum \frac{I_i - I_j}{r_{ij}^2} \quad (8)$$

Finalmente, o estado eletrotológico do átomo i é dado pela soma dos estados intrínsecos e das perturbações.

$$ES_i = I_i + \Delta I_i \quad (9)$$

Costuma-se utilizar a soma de ES_i da molécula ou de diferentes substituintes como descritor²⁵.

Para a molécula utilizada como exemplo, os valores do estado intrínseco de cada átomo são:

$$I_{(-OH)} = I_3 = [(2/2)^2 * 5 + 1] / 1 = 6; I_{(-CH_3)} = I_1 = I_6 = [(2/2)^2 * 1 + 1] / 1 = 2; \\ I_{(-CH_2-)} = I_4 = I_5 = [(2/2)^2 * 2 + 1] / 2 = 1,5; I_{(-CH<-)} = I_2 = [(2/2)^2 * 3 + 1] / 3 = 1,33; \\ r_{12} = 2; r_{13} = 3; r_{14} = 3; r_{15} = 4; r_{16} = 5; r_{23} = 2; r_{24} = 2; \\ r_{25} = 3; r_{26} = 4; r_{34} = 3; r_{35} = 4; r_{36} = 5; r_{45} = 2; r_{46} = 3; r_{56} = 2$$

A partir desses valores e dos valores de r_{ij} obtém-se a matriz, cujos elementos são $(I_i - I_j) / r_{ij}^2$:

$$\begin{bmatrix} 0 & 0,168 & -0,444 & 0,056 & 0,031 & 0 \\ -0,168 & 0 & -1,168 & -0,043 & -0,012 & -0,027 \\ 0,444 & 1,168 & 0 & 0,500 & 0,281 & 0,160 \\ -0,056 & 0,043 & -0,500 & 0 & 0 & -0,056 \\ -0,031 & 0,012 & -0,281 & 0 & 0 & -0,125 \\ 0 & 0,027 & -0,160 & 0,056 & 0,125 & 0 \end{bmatrix} \begin{matrix} -0,189 \\ -1,418 \\ 2,553 \\ -0,569 \\ -0,425 \\ 0,048 \end{matrix}$$

Na coluna da direita estão os valores de ΔI_i . A partir destes valores são obtidos os índices ES_i de acordo com a Equação 9:

$$ES_1 = 2 - 0,189 = 1,811; ES_2 = 1,33 - 1,418 = -0,088; ES_3 = 6 + 2,553 = 8,553;$$

$$ES_4 = 1,5 - 0,569 = 0,931; ES_5 = 1,5 - 0,425 = 1,075; ES_6 = 2 + 0,048 = 2,048$$

Portanto a soma dos índices eletrotológicos é $Ss = 14,33$.

Os índices topológicos utilizados neste trabalho foram calculados pelo programa Polly²⁶ e organizados em uma matriz $X(209,21)$ (matriz de variáveis independentes). Os símbolos utilizados para identificá-los estão na Tabela 1. Os valores experimentais de $\log K_{ow}$ são representados pelo vetor $y(30,1)$ das variáveis dependentes e estão entre parêntesis na última coluna da Tabela 2.

O processo geral de modelagem consiste de duas etapas: a primeira é a modelagem propriamente dita, que estabelece uma relação matemática entre X e y ; a segunda é a validação, que otimiza esta relação matemática no sentido de se obter uma melhor descrição da

propriedade em estudo. A validação cruzada, excluindo uma amostra de cada vez do conjunto de dados, foi feita para determinar o número de variáveis latentes a ser utilizado no modelo. Concluídas estas etapas, o modelo está apto para prever $\log K_{ow}$ daqueles PCBs para os quais não se dispõe de valores experimentais. Antes da modelagem, os dados foram autoescalados, *i.e.*, centrados na média e escalados pela variância¹⁷. O método de regressão utilizado foi PLS, regressão por mínimos quadrados parciais. A seleção das variáveis para a construção do modelo, foi feita por meio de uma análise hierárquica de agrupamentos (HCA)¹⁷. Os resultados desta análise são expressos na forma de um dendograma, que agrupa as variáveis de acordo com o grau de similaridade entre elas. Os programas MATLAB 5.1²⁷ e PLS_Toolbox 2.0²⁸ foram utilizados para a modelagem. O programa de quimiometria Pirouette versão 3.0²⁹ foi utilizado para a obtenção do dendograma.

RESULTADOS E DISCUSSÃO

Inicialmente foram selecionados os descritores que apresentaram uma melhor correlação com $\log K_{ow}$. A Figura 3 mostra o dendograma relativo à similaridade entre os índices topológicos e $\log K_{ow}$. Os ramos horizontais correspondem às variáveis e os números no topo da Figura indicam o índice de similaridade, numa escala que vai de **1** (identidade) a **0** (nenhuma similaridade). Os descritores que apresentaram maior similaridade com $\log K_{ow}$ foram índice de Wiener (W), índices de conectividade e soma dos índices eletrotológicos (Ss). Excetuando os índices de ramificação de ordem 4 (VC4 e SC4) que formam um conjunto à parte, pode ser observado na Figura 3, uma alta similaridade entre os descritores (variando entre 0,75 e 0,95) indicando que existe uma alta correlação

Tabela 1. Índices topológicos utilizados neste trabalho

W	W índice de Wiener
HD	H_D índice de Harary
Ss	Ss soma dos índices eletrotológicos
Sm (m=0-6)	${}^m\chi$ índice de conectividade de ordem m
SCm (m=3-6)	${}^m\chi_c$ índice de ramificação de ordem m
SPCm (m=4-6)	${}^m\chi_{pc}$ índice de caminho/ramificação de ordem m
Vm (m=0-6)	${}^m\chi^v$ índice de conectividade de valência de ordem m
VCm (m=3-6)	${}^m\chi_c^v$ índice de ramificação de valência de ordem m

Tabela 2. Descritores utilizados na modelagem e os valores previstos e experimentais* de $\log K_{ow}$

PCB	Estrutura**	W	${}^1\chi$	${}^3\chi_c$	${}^4\chi_{pc}$	${}^3\chi_c^v$	$\log K_{ow}$	PCB	Estrutura**	W	${}^1\chi$	${}^3\chi_c$	${}^4\chi_{pc}$	${}^3\chi_c^v$	$\log K_{ow}$
1	(2)	240	6,38	0,54	1,44	0,32	4,62(4,56)	54	(2,2',6,6')	394	7,61	1,17	2,77	0,78	6,17(5,94)
2	(3)	246	6,36	0,62	1,29	0,36	4,67(4,72)	61	(2,3,4,5)	412	7,61	1,12	3,23	0,74	6,24(6,39)
3	(4)	252	6,36	0,62	1,35	0,36	4,69(4,69)	70	(2,3',4',5)	426	7,58	1,30	2,88	0,83	6,35(6,39)
4	(2,2')	287	6,79	0,74	1,94	0,47	5,14(5,02)	77	(3,3',4,4')	440	7,58	1,28	3,12	0,82	6,40(6,52)
5	(2,3)	291	6,79	0,73	2,07	0,46	5,16(5,20)	80	(3,3',5,5')	426	7,54	1,49	2,19	0,92	6,37(6,58)
7	(2,4)	298	6,77	0,83	1,78	0,51	5,19(5,15)	87	(2,2',3,4,5')	486	8,00	1,42	3,53	0,94	6,85(6,58)
9	(2,5)	293	6,77	0,83	1,79	0,51	5,18(5,18)	101	(2,2',4,5,5')	488	7,99	1,50	3,33	0,99	6,89(6,85)
11	(3,3')	301	6,75	0,91	1,64	0,55	5,24(5,34)	116	(2,3,4,5,6)	470	8,04	1,24	3,81	0,86	6,73(6,85)
15	(4,4')	315	6,75	0,91	1,76	0,55	5,29(5,28)	128	(2,2',3,3',4,4')	569	8,43	1,51	4,41	1,04	7,41(7,44)
18	(2,2',5)	346	7,18	1,03	2,30	0,66	5,71(5,64)	153	(2,2',4,4',5,5')	573	8,40	1,69	4,00	1,13	7,49(7,44)
28	(2,4,4')	368	7,17	1,12	2,19	0,70	5,80(5,74)	185	(2,2',3,4,5,5',6)	628	8,84	1,73	4,67	1,20	7,91(7,93)
29	(2,4,5)	354	7,18	1,01	2,47	0,65	5,74(5,77)	194	(2,2',3,3',4,4',5,5')	738	9,25	1,90	5,53	1,32	8,58(8,68)
31	(2,4',5)	362	7,17	1,12	2,20	0,7	5,79(5,77)	202	(2,2',3,3',5,5',6,6')	702	9,25	1,94	5,19	1,35	8,48(8,42)
37	(3,4,4')	374	7,17	1,09	2,44	0,68	5,84(5,90)	206	(2,2',3,3',4,4',5,5',6)	820	9,68	2,02	6,11	1,43	9,11(9,14)
52	(2,2',5,5')	412	7,58	1,32	2,65	0,85	6,30(6,26)	209	(2,2',3,3',4,4',5,5',6,6')	907	10,11	2,15	6,70	1,54	9,64(9,60)

*Entre parêntesis estão os valores experimentais¹¹ de $\log K_{ow}$; ** notação conforme Figura 1

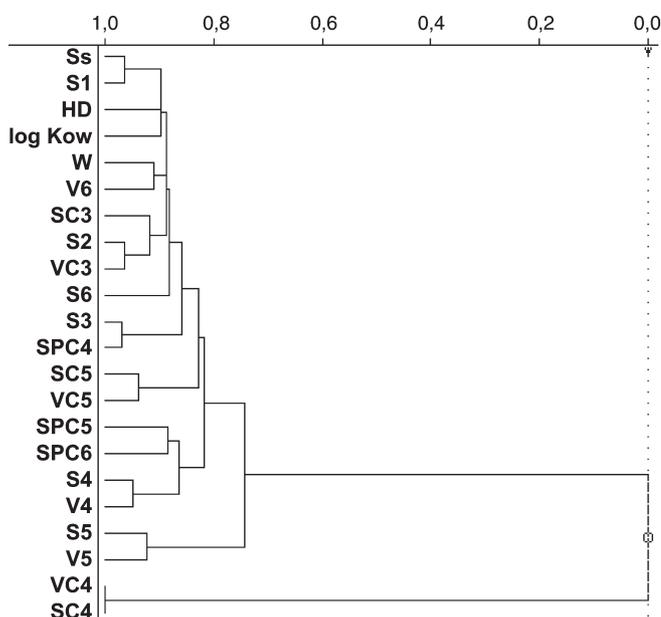


Figura 3. Dendrograma obtido da análise de agrupamento hierárquico, mostrando a similaridade entre $\log K_{ow}$ e os índices topológicos

entre os mesmos. Entretanto este fato não se constitui num problema para a construção de modelos de regressão, uma vez que o método utilizado é o de mínimos quadrados parciais (PLS) e não o método tradicional de regressão linear múltipla (MLR)¹⁷. No método PLS, as variáveis correlacionadas são agrupadas para formar as variáveis latentes contornando assim os problemas de colinearidade, inclusive permitindo a modelagem mesmo quando se tem um reduzido número de compostos (menor que o número de descritores).

A Tabela 2 contém todos os valores dos descritores que foram selecionados, os quais apresentam um comportamento crescente com $\log K_{ow}$, ou seja, são diretamente proporcionais (Figura 4). São eles: (i) o índice de Wiener (W). De acordo com Platt,³⁰ este índice está relacionado com o volume molecular. Mais precisamente, W representa o quão compacta é a molécula. Quanto mais compactada, menor o valor de W e, portanto, menor o valor de $\log K_{ow}$; (ii) o índice de conectividade de primeira ordem (${}^1\chi$). Na literatura, o índice de conectividade de primeira ordem é encontrado em diversos modelos, relacionado com o volume molecular³¹, com a área³² e com o tempo de retenção de benzenos clorados³³; (iii) o índice de ramificação de ordem 3 (${}^3\chi_c$). Este índice é mais sensível às variações estruturais do que ${}^1\chi$, e descreve principalmente o grau de ramificação da molécula. ${}^3\chi_c$ foi utilizado com sucesso, por exemplo, na estimativa de pressão de vapor de compostos de interesse ambiental³⁴; (iv) o índice de quarta ordem de caminho/ramificação (${}^4\chi_{pc}$). Este índice tem sido relacionado à substituição por átomos de cloro em diversos compostos³⁵. Também tem sido observado que, em uma mesma clas-

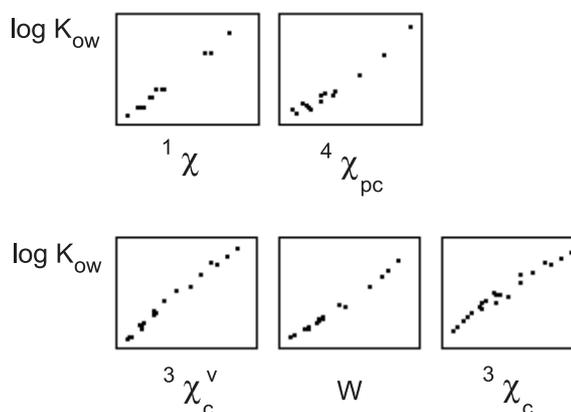


Figura 4. Correlação entre os descritores selecionados e $\log K_{ow}$

se de isômeros, o valor de ${}^4\chi_{pc}$ aumenta com a proximidade dos átomos de cloro³³; (v) por fim, o índice de ramificação de valência de ordem 3 (${}^3\chi_c^v$). O índice ${}^3\chi_c^v$ contém não apenas informações sobre a ramificação da molécula, mas também sobre os elétrons de valência. Então ${}^3\chi_c^v$ pode refletir alguma contribuição eletrônica, decorrente da presença de átomos de cloro. Em virtude da metodologia usada para a construção dos modelos ser PLS, o fato deles serem ou não altamente correlacionados entre si não se constitui numa desvantagem, mas pelo contrário, em vantagem. Como os descritores são utilizados simultaneamente na modelagem, toda a informação relevante intrínseca de cada um é incluída no modelo, enquanto que as informações redundantes são eliminadas.

Foram construídos modelos utilizando três descritores. Os dados estatísticos foram comparados para a escolha do melhor modelo. De acordo com Wold³⁶, as condições estatísticas inerentes a um modelo de boa qualidade incluem um desvio padrão residual (s) < 0,30, o quadrado do coeficiente de correlação (r^2) > 0,90, e soma dos quadrados residuais (PRESS) < 0,40. O coeficiente de correlação de validação cruzada (Q^2) > 0,90 é indicação de um excelente modelo³⁷. Também foram empregados o desvio padrão dos erros de previsão (S_{PRESS}), e o teste de Fischer (F). Neste teste, o valor de F calculado é comparado com o valor tabelado. Uma vez que o valor de F calculado seja maior que o valor de F tabelado, conclui-se que o modelo obtido não é resultado de uma correlação aleatória. Com base nestes critérios, e nos resultados estatísticos para os modelos indicados na Tabela 3, podemos afirmar que todas as combinações apresentadas fornecem modelos de excelente qualidade. Todos eles apresentaram bons coeficientes de regressão com duas variáveis latentes. O modelo escolhido, (4), foi aquele que previu K_{ow} para todas as moléculas com o menor resíduo na etapa de validação, e que forneceu o maior valor de F calculado.

No modelo 4, a primeira variável latente (LV) acumula 96,63% da variância total, e é o resultado da contribuição de todas as variá-

Tabela 3. Dados estatísticos para os modelos de $\log K_{ow}$

Modelo	Descritor	r	PRESS	S_{PRESS}	s	Q^2	F-calculado
1	$W, {}^1\chi, {}^4\chi_{pc}$	0,995	0,449	0,128	0,114	0,992	$3,14 \times 10^3$
2	$W, {}^3\chi_c, {}^3\chi_c^v$	0,997	0,361	0,112	0,114	0,993	$3,88 \times 10^3$
3	${}^3\chi_c, {}^1\chi, {}^4\chi_{pc}$	0,997	0,361	0,116	0,116	0,993	$3,87 \times 10^3$
4	${}^3\chi_c, {}^4\chi_{pc}, {}^3\chi_c^v$	0,997	0,244	0,095	0,100	0,995	$5,75 \times 10^3$

veis. A segunda, descreve uma variância de 3,36% dos dados originais e apresenta uma contribuição significativa da variável ${}^4\chi_{pc}$. A terceira variável latente contribui com apenas 0,01%. O valor de PRESS é menor quando se utilizam três LVs, mas a contribuição da terceira LV é muito pequena (menor que 1,0% para dados auto-escalados). Então o modelo foi construído com duas LVs e os resultados obtidos estão na Tabela 4. Na Equação 11 está a equação final obtida para o cálculo de K_{ow} a partir de três descritores topológicos juntamente com os resultados estatísticos obtidos na etapa de validação do modelo, em que n_a é o número de amostras utilizadas na modelagem. Na Figura 5 são mostrados os valores previstos para todas as moléculas que participaram da modelagem *versus* os valores experimentais. Estes valores previstos estão também incluídos na última coluna da Tabela 2.

$$\log K_{ow} = 3,35 + 0,90 ({}^3\chi_c) + 0,49 ({}^4\chi_{pc}) + 1,25 ({}^3\chi_v) \quad (11)$$

$$r^2 = 0,9944 \quad s = 0,09 \quad Q^2 = 0,9943$$

$$PRESS = 0,244 \quad S_{PRESS} = 0,090 \quad F = 5,75 \times 10^3 \quad F_{2,27} = 5,49 \quad n_a = 30$$

Tabela 4. Porcentagem da variância acumulada em cada variável latente pelo modelo PLS para $\log K_{ow}$

#LV	Bloco X		Bloco Y	
	Esta LV	Total	Esta LV	Total
1	96,63	96,63	99,43	99,43
2	3,36	99,99	0,10	99,53

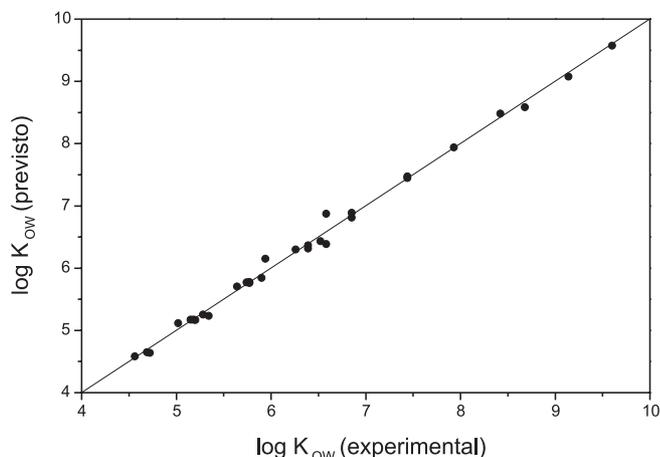


Figura 5. Gráfico dos valores previstos e experimentais de $\log K_{ow}$

Uma maneira de avaliar o desempenho do modelo, seria comparar aos resultados experimentais com os previstos por meio do coeficiente de variabilidade $CV = \frac{\text{desvio padrão dos resíduos}}{\text{média dos valores experimentais}} \times 100$, onde o resíduo é definido como $(\log K_{ow}^{\text{experimental}} - \log K_{ow}^{\text{previsto}})$. O ideal é que o coeficiente de variabilidade seja uma pequena fração do valor médio experimental. O valor médio dos dados experimentais é 6,37 e o desvio padrão dos resíduos obtido foi de 0,101, resultando num coeficiente de variabilidade igual a 1,58 (apenas 1,58% do valor médio dos dados experimentais), que é um valor bastante baixo. Se o desvio padrão dos resíduos for da mesma ordem de magnitude do desvio padrão dos dados experimentais, indica que a Equação 11 não prevê bem os valores experimentais. Idealmente, o valor da razão $\frac{\text{desvio padrão dos valores experimentais}}{\text{desvio padrão dos resíduos}}$ deve ser

igual ou maior que 10,0. No presente trabalho, o desvio padrão dos resultados experimentais é de 1,34 dando uma razão de 13,34, o que nos leva a concluir que o modelo obtido é de excelente qualidade e prevê bem os resultados experimentais.

Uma vez obtida e validada a função matemática que relaciona os descritores com a propriedade em questão, esta relação (modelo) é utilizada para previsão. Os resultados previstos para os outros 179 PCBs podem ser obtidos com os autores.

Na literatura existem alguns resultados de modelos para prever $\log K_{ow}$ de PCBs, os quais podem ser comparados com os resultados obtidos neste trabalho. Hawker e colaboradores³⁸ correlacionaram $\log K_{ow}$ com o $\log \alpha$ (tempo de retenção relativo), o resultado obtido apresentou $r^2 = 0,90$, $s = 0,33$ com $n_a = 30$. Gramatica e colaboradores³⁹ correlacionaram $\log K_{ow}$ de PCBs com descritores WHIM relacionados ao tamanho molecular, na modelagem foi obtido $r^2 = 0,95$ com $n_a = 139$. Hansen e colaboradores⁴⁰ correlacionaram $\log K_{ow}$ com a área da superfície molecular, obtendo $r^2 = 0,96$ com $n_a = 53$, sendo que os PCBs foram divididos em dois grupos de acordo com o número de átomos de cloro na posição *orto*, e foi obtida uma equação para cada grupo. Também podemos verificar que a média residual do modelo 4 (0,063) é menor que a média residual obtida pelo programa CLOGP (0,141)⁴¹, o qual é adequado para PCBs com um máximo de cinco átomos de cloro na molécula.

Analisando a Equação 11, observa-se que os coeficientes dos índices topológicos utilizados são todos positivos, como esperado (Figura 4). Como mencionado anteriormente, ${}^3\chi_c$ aumenta com o aumento da ramificação na molécula, e no caso dos PCBs esta ramificação é diretamente relacionada à inclusão de átomo de cloro. Por exemplo, a molécula 18 (2,2',5-CB) com $\log K_{ow} = 5,64$ é mais lipofílica que a molécula 4 (2,2'-CB) que possui $\log K_{ow} = 5,02$. O índice ${}^3\chi_c^v$ foi introduzido por Hall e Randić⁴² para diferenciar os átomos de carbono dos heteroátomos. O índice ${}^3\chi_c$ é limitado aos elétrons σ , enquanto que o índice ${}^3\chi_c^v$ inclui todos os elétrons de valência da molécula. Quando se calcula o delta de valência, são considerados os elétrons que podem fazer as ligações π e σ . Os elétrons π são mais expostos e mais interativos que os elétrons σ , sendo mais significativos em interações intermoleculares não covalentes. O átomo de cloro possui elétrons no orbital p que podem fazer ligações tipo π . Então, ${}^3\chi_c^v$ pode estar relacionado à maior disponibilidade de elétrons que podem fazer ligações π devido ao aumento de átomos de cloro na molécula. O índice ${}^4\chi_{pc}$ contém informação sobre a distância entre os átomos de cloro na molécula, quanto maior a proximidade entre os átomos de cloro, maior o valor de ${}^4\chi_{pc}$ e, portanto, de $\log K_{ow}$, por exemplo $\log K_{ow}$ do PCB 5 (2,3-CB) é maior que $\log K_{ow}$ do PCB 7 (2,4-CB). Nenhum dos descritores considera a posição do átomo de cloro no anel, apenas a disposição dos mesmos. Mas analisando a Tabela 2 é possível notar a influência da posição em que o átomo de cloro está no anel. Por exemplo, o $\log K_{ow}$ do PCB 4 com os dois átomos de cloro na posição *orto* (2,2'-CB), apresenta menor valor (5,02) que o $\log K_{ow}$ do PCB 11 (5,34) no qual os dois átomos de cloro estão na posição *para* (3,3'-CB). $\log K_{ow}$ do PCB 28 (5,74) é menor que $\log K_{ow}$ do PCB 37 (5,90), sendo que a única diferença é a mudança do átomo de cloro da posição *orto* (2,4,4'-CB) para a posição *para* (3,4,4'-CB). Até agora foram comparadas as influências dos átomos de cloro nas posições *orto* e *para*, mas é possível analisar a influência da posição *meta*. $\log K_{ow}$ do PCB 15 (5,28) o qual apresenta os átomos de cloro na posição *meta* (4,4'-CB), é menor que $\log K_{ow}$ do PCB 11 (5,34), no qual os átomos de cloro estão na posição *para* (3,3'-CB). A importância do átomo de cloro na posição *para* fica clara quando se observa que $\log K_{ow}$ do PCB 80 (6,58), com quatro átomos de cloro nas posições *para* (3,3',5,5'-CB), apresenta o mesmo valor que $\log K_{ow}$ do PCB 87 que, apesar de apresentar um átomo de cloro a mais,

possui apenas dois átomos de cloro na posição *para* (2,2',3,4,5'-CB). De fato, átomos de cloro na posição *para* parecem contribuir mais para a lipofilicidade dos PCBs do que átomos de cloro na posição *orto* e *meta*. Sendo a ordem de importância para o aumento da lipofilicidade a seguinte: *para*, *meta* e *orto*.

CONCLUSÕES

A aplicação apresentada neste trabalho mostra como é possível obter os valores de $\log K_{ow}$, uma importante propriedade dos PCBs, para todos os congêneros, utilizando apenas índices topológicos. De uma maneira geral, todos os índices estão relacionados não apenas ao número de átomos de cloro na molécula, mas também à posição e proximidade destes substituintes. Uma das vantagens em se utilizar estes índices está no fato de não ser necessário o conhecimento prévio da estrutura tridimensional da molécula o que, em alguns casos, é difícil de se obter. Destaca-se também a facilidade com que estes índices podem ser calculados, sua natureza não empírica e a sua capacidade em descrever propriedades estruturais locais (por exemplo $^1\chi$) ou globais (por exemplo W), assim como propriedade específica relacionada à variação de um determinado substituinte. De um modo geral, a partir de modelos obtidos por meio de índices topológicos é possível prever, por exemplo, o quanto um composto, ainda não sintetizado, pode ser perigoso para o meio ambiente, ou se uma droga pode ou não ser eficiente.

Os resultados obtidos neste trabalho, mostram que o aumento de átomos de cloro dos PCBs pode ser responsável pelo aumento de ramificações e de elétrons π na molécula, e a proximidade entre estes átomos, seria então responsável pelo aumento de $\log K_{ow}$.

AGRADECIMENTOS

Agradecemos ao Prof. Dr. S. Basak e à Dra. D. Mills pela utilização do programa POLLY e ao apoio das agências financiadoras, CNPq (L. R. Cirino) e FAPESP (M. M. C. Ferreira).

REFERÊNCIAS

- Rachdawong, P.; Christensen, E. R.; *Environ. Sci. Technol.* **1997**, *31*, 2686.
- Ballschmitter, K.; Zell, M.; *Fresenius' J. Anal. Chem.* **1980**, *302*, 20.
- Hutzinger, O.; Safe, S.; Zitko V.; *The Chemistry of PCBs*, CRC Press: Boca Raton, FL, 1974.
- Giesy, J. P.; Kannan, K.; *Crit. Rev. Toxicol.* **1998**, *28*, 511.
- Mullin, M. D.; Pochini, C. M.; McCrindle, S.; Romkes, M.; Safe, S. H.; Safe, L. M.; *Environ. Sci. Technol.* **1984**, *18*, 468.
- Shiu, W. Y.; Mackay, D.; *J. Phys. Chem. Ref. Data* **1986**, *15*, 911.
- Watari, H.; Tanaka, M.; Suzuki, N.; *Anal. Chem.* **1982**, *54*, 702.
- Dearden, J. C.; *Environ. Sci. Technol.* **1985**, *61*, 203; Mallholt, H.; Peters, R. H.; *Environ. Sci. Technol.* **1988**, *22*, 1479; Ferreira, M. M. C.; *Chemosphere* **2001**, *44*, 125.
- Rapaport, R. A.; Eisenreich, S. J.; *Environ. Sci. Technol.* **1984**, *18*, 163.
- Miller, M. M.; Ghodbane, S.; Wasik, S. P.; Tewari, Y. B.; Martire, D. E.; *J. Chem. Eng. Data* **1984**, *29*, 184.
- Mackay, D.; Shiu, W. Y.; Ma, K. C.; *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemical*, Lewis Publishers: Boca Raton, FL, 1993, vol. 3.
- Leo, A.; Jow, P. Y. C.; Silipo, C.; Hansch, C.; *J. Med. Chem.* **1975**, *18*, 865; Nys, G. G.; Rekker, R. F.; *Eur. J. Med. Chem.* **1974**, *9*, 361; Suzuki, T.; Kudo, Y.; *J. Comput.-Aided Mol. Des.* **1990**, *4*, 155.
- Broto, P.; *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163; Broto, P.; Moreau, G.; Vanduycke, C.; *Eur. J. Med. Chem.* **1984**, *19*, 71; Ghose, A. K.; Crippen, G. M.; *J. Comput. Chem.* **1986**, *4*, 565.
- Klopman, G.; Iroff, L. D.; *J. Comput. Chem.* **1981**, *2*, 157; Bodor, N.; Huang, M.-J.; *J. Comput. Chem.* **1991**, *12*, 1182.
- Hansch, C.; Fujita, T.; *J. Am. Chem. Soc.* **1964**, *86*, 1616.
- Franke, R.; *Theoretical Drug Design Methods*, Elsevier: New York, 1984; Benigni, R.; Richard, M.; *Methods* **1998**, *14*, 264; Schultz, T. W.; Seward, J. R.; *Sci. Total Environ.* **2000**, *249*, 73.
- Ferreira, M. M. C.; *J. Braz. Chem. Soc.* **2002**, *13*, 742; Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O.; *Quim. Nova* **1999**, *22*, 724; Beebe, K. R.; Pell, P. J.; Seasholtz, M. B.; *Chemometrics: A Practical Guide*, Wiley: New York, 1998.
- Basak, S. C.; *Med. Sci. Res.* **1987**, *15*, 605; Raychaudhury, C.; Ray, S. K.; Ghosh, J. J.; Roy, A. B.; Basak, S. C.; *J. Comput. Chem.* **1984**, *5*, 581.
- Trinajstić, N.; *Chemical Graph Theory*, CRC: Boca Raton, FL, 1992, cap. 3; Neves, P. J.; Costa, J. B. N.; Ndiya, P. M.; Carneiro, J. W. M.; *Quim. Nova* **1998**, *21*, 1.
- Kier, L. B.; Hall, L. H.; *Molecular Connectivity in Structure-Activity Analysis*, John Wiley & Sons: New York, 1986.
- Wiener, H.; *J. Am. Chem. Soc.* **1947**, *69*, 2636.
- Randić, M.; *J. Am. Chem. Soc.* **1975**, *97*, 6609.
- Plavšić, D.; Nikolaić, S.; Trinajstić, N.; Mihalić, Z.; *J. Math. Chem.* **1993**, *12*, 235.
- Kier, L. B.; Hall, L. H.; *J. Math. Chem.* **1991**, *7*, 229; Hall, L. H.; Mohney, B. K.; Kier, L. B.; *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76.
- Sullivan, J. J.; Jones, A. D.; Tanji, K. K.; *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1113.
- Basak, S.; Harriss, D.; Magnuson, V.; *POLLY*, Natural Resources Research Institute: University of Minnesota, Duluth, 1988.
- Moler, C.; Little, J.; Kleiman, S.; Bangert, S.; *MATLAB™*, The Math Works Inc.: Natick, Ma, 1997.
- Wise, B. M.; Gallagher, N. B.; *PLS Toolbox*, Eigenvector Technologies: West Richland, WA, 1995.
- Pirouette, Multivariate Data Analysis for IBM PC Systems*; Infometrix Inc.: Woodinville, WA, 2001.
- Platt, J. R.; *J. Phys. Chem.* **1952**, *56*, 328.
- Feng, L.; Zhang, W.; Wang, L.; Wei, L.; *Chemosphere* **1996**, *32*, 2505.
- Sabljić, A.; *Environ. Sci. Technol.* **1987**, *21*, 358.
- Sabljić, A.; *J. Chromatogr.* **1985**, *319*, 1.
- Sacan, M. T. S.; Balcioglu, I. A.; *Chemosphere* **1998**, *36*, 451.
- Kier, L. B.; *J. Pharm. Sci.* **1980**, *69*, 1034; Sabljic, A.; Protic, M.; *Mol. Pharmacol.* **1983**, *23*, 213; Sabljic, A.; *J. Chromatogr.* **1984**, *314*, 1.
- Wold, S.; *Quant. Struct.-Act. Relat.* **1991**, *10*, 191.
- Boyd, R. J.; Boyd, S. L.; *J. Am. Chem. Soc.* **1992**, *114*, 1652.
- Hawker, D. W.; Connell, D. W.; *Environ. Sci. Technol.* **1988**, *22*, 382.
- Gramatica, P.; Navas, N.; Todeschini, R.; *Chem. Intelligent Lab. Syst.* **1998**, *40*, 53.
- Hansen, B. J.; Paya-Perez, A. B.; Rahman, M.; Larsen, B. R.; *Chemosphere* **1999**, *39*, 2209.
- Chessells, M.; Hawker, D. W.; de Bruijn, J.; Hermens, J. L. M.; *Chemosphere* **1990**, *18*, 457.
- Kier, L. B.; Hall, L. H.; *J. Pharm. Sci.* **1983**, *72*, 1170; Hall, L. H.; Kier, L. B.; *J. Mol. Graph. Mod.* **2001**, *20*, 4.