

**TOP - UM PROGRAMA DE CÁLCULO DE DESCRITORES TOPOLÓGICOS PARA USO EM CORRELAÇÕES ENTRE ESTRUTURA E ATIVIDADE<sup>1</sup>**Porfírio Jesus das Neves<sup>a</sup>, João Batista Neves da Costa<sup>b</sup> e Papa Matar Ndiyae

Departamento de Química - Instituto de Ciências Exatas - UFRRJ

José Walkimar de Mesquita Carneiro<sup>c</sup>

Departamento de Química Geral e Inorgânica - Instituto de Química - UFF

Recebido em 24/7/97; aceito em 13/4/98

**TOP - A SOFTWARE FOR CALCULATION OF TOPOLOGICAL DESCRIPTORS TO BE USED IN STRUCTURE - ACTIVITY RELATIONSHIPS. A computational interface, using sybyl format as an input file, was created in order to calculate connectivity indexes. After generation and classification of all substructures, which derive from the molecular structure, this interface calculates all possible orders from zero up to the maximum number of bonds in the molecule. Other topological indexes such as Wiener and Schultz indexes can also be calculated.**

**Keywords: topological index; connectivity index; QSAR.**

**INTRODUÇÃO**

Métodos para obtenção de correlações quantitativas entre estrutura molecular e atividade (QSAR) e/ou propriedades (QSPR) baseiam-se em três hipóteses básicas. A primeira é que a estrutura de uma molécula, quer nos aspectos estéricos quer nos aspectos eletrônicos, deve conter as informações responsáveis por suas propriedades físicas, químicas ou biológicas. A segunda hipótese é que estas informações podem ser representadas de forma quantitativa por parâmetros numéricos para a atividade e/ou propriedade. Por fim, a possível correlação entre a estrutura molecular e a atividade/propriedade que se quer otimizar deve ser expressa em termos de relações matemáticas simples que permitam ao mesmo tempo fazer previsões de atividade/propriedade para sistemas análogos e, mais importante, permitam uma interpretação do processo em termos químicos<sup>2</sup>. A essência dos métodos QSAR é, portanto, transformar a estrutura química de um composto em uma série de descritores numéricos que representem as características mais relevantes para uma dada atividade/propriedade e, adicionalmente, estabelecer relações quantitativas entre os descritores e a atividade/propriedade que se quer otimizar.

A possibilidade de que a atividade e/ou propriedade seja uma função da estrutura molecular é uma hipótese intuitiva e vastamente discutida<sup>3</sup>. A segunda hipótese acima, qual seja, que a estrutura de uma molécula seja representada quantitativamente por descritores é menos intuitiva e também de ampla discussão<sup>4</sup>. Existe uma variedade de procedimentos para representar estrutura em termos de descritores, os quais, em maior ou menor grau, contêm as propriedades relevantes da estrutura molecular. Dentre as várias metodologias disponíveis para representação de estrutura molecular em termos de parâmetros numéricos, uma que merece destaque é a que procura representar uma molécula em termos de informações contidas em sua estrutura bidimensional, essencialmente aquelas relacionadas à conectividade atômica que dão origem à estrutura molecular. Estes parâmetros são conhecidos de uma forma geral como índices topológicos<sup>5</sup>.

Índices topológicos têm encontrado considerável sucesso na previsão de uma grande variedade de propriedades físicas, químicas ou biológicas<sup>5</sup>. Propriedades tais como calor de

vaporização<sup>5</sup>, calor de formação<sup>5</sup>, ponto de ebulição<sup>6,7</sup>, refração molar<sup>6</sup>, solubilidade<sup>7</sup>, densidade<sup>8</sup>, coeficiente de partição<sup>9</sup>, polaridade, tempo de retenção em cromatografia<sup>10,11</sup>, etc., têm sido correlacionadas com descritores derivados dos índices topológicos. Estes índices também vêm sendo empregados na obtenção de correlação com diferentes atividades biológicas, podendo-se citar anestésicas<sup>12,13,14</sup>, narcóticas<sup>4</sup>, halucinogênicas<sup>15</sup>, inibições enzimáticas<sup>5,16</sup>, bromatológicas<sup>17</sup>, analgésicas<sup>18</sup>, anticonvulsivantes<sup>19</sup>, etc.

O cálculo de uma série de descritores topológicos para um conjunto de moléculas é, do ponto de vista matemático, consideravelmente simples, contudo, pode tornar-se uma tarefa laboriosa e com grande chance de erro se aplicada manualmente a um grande número de moléculas, principalmente quando estas possuem estrutura complexa. O emprego de metodologias computacionais é portanto recomendável não apenas pela redução no tempo necessário para tratar um determinado número de moléculas, mas principalmente, por evitar a chance de erro na obtenção dos descritores. Embora algumas metodologias computacionais para o cálculo de descritores topológicos tenham sido descritas na literatura, estas não são gerais e nem sempre são de fácil acesso<sup>20,21</sup>. O presente trabalho apresenta uma metodologia simples, de fácil emprego pelo usuário e de aplicação generalizada para cálculo de descritores topológicos, principalmente aqueles desenvolvidos por Kier e Hall<sup>4,5</sup>, conhecidos como índices de conectividade molecular.

**TEORIA DOS ÍNDICES TOPOLÓGICOS**

Os índices topológicos baseados na conectividade molecular são de três tipos: índices chi de conectividade molecular,  ${}^m\chi$ , que caracterizam atributos estruturais da molécula; índices kappa de forma molecular,  ${}^m\kappa$ , e os valores de equivalência topológica T, que caracterizam átomos e grupos no esqueleto molecular e que são usados essencialmente para determinar átomos quimicamente equivalentes dentro de uma molécula<sup>4,5</sup>. Outros tipos de índices também merecem destaque, podendo-se citar, por exemplo, os índices eletrotológicos<sup>22</sup>.

A análise da estrutura molecular do ponto de vista da conectividade começa pela adoção de uma representação apropriada para esta estrutura. Esta representação é baseada no esqueleto molecular, o qual contém a rede de ligações químicas, incluindo os átomos e as conexões entre eles. Tal representação é chamada de gráfico molecular. Este por sua vez é

<sup>a</sup> porfirio@ufrj.br; <sup>b</sup> dacosta@ufrj.br; <sup>c</sup> gqjose@vm.uff.br

constituído de vértices, representados pelos átomos, e lados, representados pelas ligações. A série de átomos e conexões, no gráfico molecular, contém informação estrutural a qual deve ser transformada em um índice numérico que pode ser usado para representar a estrutura molecular. Os índices devem contemplar principalmente os elementos que são variáveis com a estrutura.

O menor nível de informação estrutural em um gráfico molecular é simplesmente o número de vértices. Obviamente este é um índice com baixíssimo conteúdo de informação. Outros índices com maior grau de informação podem ser abstraídos a partir do gráfico molecular. Índices que representam simplesmente o número de lados (índice de Wiener) ou o número de pares de lados adjacentes (índice de Platt<sup>23</sup>) em um gráfico molecular são exemplos de índices simples mas que podem fornecer interessantes correlações com uma série de propriedades. O avanço natural é no sentido de desenvolver índices que incluam maior quantidade de informação estrutural. Neste sentido, Randic introduziu o conceito de grau de vértice,  $\delta$ , o qual, para um dado vértice, é definido como o número de vértices vizinhos<sup>24</sup>. Por exemplo, o grupo metileno possui apenas um vizinho, portanto  $\delta = 1$ ; um grupo metileno tem dois vizinhos,  $\delta = 2$ , e daí por diante. Randic também propôs um peso para o lado entre os vértices  $i$  e  $j$ , definindo-o por  $(\delta_i \cdot \delta_j)^{-1/2}$ . A partir deste ponto pode-se definir um índice de ramificação como a soma dos pesos dos lados, para todos os lados em um gráfico molecular,  $(\sum(\delta_i \cdot \delta_j)^{-1/2})$ .

Expressando matematicamente, cada lado entre os vértices  $i$  e  $j$  é caracterizado por um número definido como:

$$C_{ij} = (\delta_i \cdot \delta_j)^{-1/2} \quad (1)$$

onde  $C_{ij}$  é o peso entre os vértices  $i$  e  $j$  e  $\delta_i$  é o grau do vértice  $i$ .

O índice de ramificação ou de conectividade molecular,  $\chi$  é obtido fazendo-se a soma de todos os  $C_{ij}$ .

$$\chi = \sum C_{ij} = \sum (\delta_i \cdot \delta_j)^{-1/2} \quad (2)$$

A generalização do índice de Randic foi feita por Kier e Hall. Estes definiram o índice  $\chi$  em várias ordens,  ${}^m\chi$ , em função do número de lados considerados em cada caso, como pode ser mostrado pelas equações a seguir:

$${}^0\chi = \sum (\delta_i)^{-1/2} \quad (3)$$

$${}^1\chi = \sum (\delta_i \cdot \delta_j)^{-1/2} \quad (4)$$

$${}^2\chi = \sum (\delta_i \cdot \delta_j \cdot \delta_k)^{-1/2} \quad (5)$$

Segundo Kier e Hall o índice  $\chi$ , em cada ordem, representa uma característica distinta da estrutura molecular. Para ordens maiores que 2 foram definidos novos índices de conectividade que levam em consideração os tipos de subestruturas consideradas. Desta forma, pode-se ter sequências do tipo linha( $\chi_L$ ), ramo( $\chi_R$ ), linha/ramo( $\chi_{LR}$ ) ou anel( $\chi_A$ ), dependendo da molécula apresentar ramificações ou ciclos, como exemplificadas no esquema 1.

Os índices de conectividades, como definidos acima, não conseguem distinguir entre diferentes átomos nem entre dife-

	Linha(L)	Ramo(R)	Linha/Ramo	Anel(A)
Ordem 5				
Ordem 6				

Esquema 1. Exemplos típicos de subestruturas.

rentes hibridizações. Por exemplo, o *p*-nitrotolueno (a) e o limoneno (b) apresentam o mesmo gráfico molecular (c) e portanto índices de conectividades iguais (Figura 1).

Para corrigir estas deficiências Kier e Hall sugeriram a definição do grau de vértice ( $\delta$ ) não mais em função do número

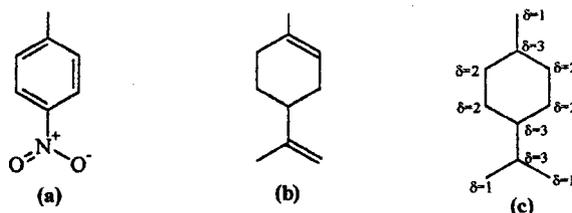


Figura 1. Estruturas (a e b) e gráfico molecular (c).

de lados vizinhos ao vértice mas em função da valência do átomo representado pelo vértice<sup>25</sup>. Desta forma o grau de vértice é definido por:

$$d_i^v = \frac{Z_i^v - h_i}{Z_i - Z_i^v - 1} \quad (6)$$

onde:

$Z_i^v$  = número de elétrons de valência do átomo  $i$ ;

$h_i$  = número de átomos de hidrogênios ligados ao átomo  $i$ ; e

$Z_i$  = número atômico do átomo  $i$ .

A relação acima leva a uma definição equivalente à anterior para carbonos saturados, mas distingue claramente a presença de insaturações ou de heteroátomos (Figura 2).

Com base no grau de vértice  $\delta^v$ , calcula-se o índice de conectividade de valência de várias ordens,  ${}^m\chi^v$ , de forma similar à definida anteriormente, conforme mostrados nas equações 3, 4 e 5.

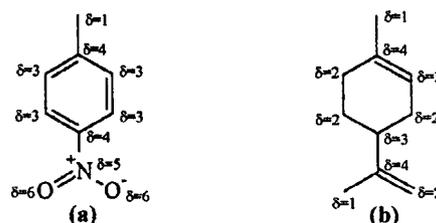


Figura 2. Graus de vértice ( $\delta^v$ ) das estruturas a e b corrigidos.

As relações expressas acima definem o cálculo dos índices de conectividade conforme desenvolvido por Kier e Hall<sup>7,8</sup>. No presente trabalho outros índices, o índice de Wiener e o índice de Schultz<sup>26</sup> foram também incluídos. Estes são calculados de forma simples. O índice de Wiener é definido como a soma das menores distâncias entre todos os vértices em um gráfico molecular e é calculado a partir da matriz de distâncias como a soma de todos os elementos  $a_{ij}/2$ .

O índice de Schultz (Molecular Topological Index-MTI) é definido como:

$$MTI = \sum_{i=1}^N e_i$$

onde,  $e_i$  ( $i = 1, 2, \dots, N$ ) são os elementos da matriz de ordem  $N$   $v.[A + D] = [e_1, e_2, \dots, e_N]$ , onde  $A$  é a matriz adjacência ( $N \times N$ ),  $D$  é a matriz distância ( $N \times N$ ) e  $v$  a matriz grau de vértice ( $1 \times N$ ).  $v$  é obtida pela soma das colunas ou linhas da matriz adjacência.



Em uma etapa subsequente cada subestrutura é classificada como sendo do tipo linha ( $\chi_L$ ), ramo ( $\chi_R$ ), linha-ramo ( $\chi_{LR}$ ) e anel ( $\chi_A$ ).

Nº de ligações			
1	2	3	4
(1-2)	(1-2)(1-6) L	(1-2)(1-6)(1-2)* (1-2)(1-6)(1-7) R	(1-2)(1-6)(1-7)(1-7)* (1-2)(1-6)(1-7)(1-2)* (1-2)(1-6)(1-7)(5-6) L/R (1-2)(1-6)(1-7)(1-2)* (1-2)(1-6)(1-7)(1-6)* (1-2)(1-6)(5-6)(1-2)* (1-2)(1-6)(5-6)(1-7)** (1-2)(1-6)(5-6)(5-6)* (1-2)(1-6)(5-6)(1-6)* (1-2)(1-6)(5-6)(4-5) L
	(1-2)(1-7) L	(1-2)(1-7)(1-2)* (1-2)(1-7)(1-6)**	
	(1-2)(2-3) L	(1-2)(2-3)(1-2)* (1-2)(2-3)(3-4) L	(1-2)(2-3)(3-4)(1-2)* (1-2)(2-3)(3-4)(3-4)* (1-2)(2-3)(3-4)(2-3)* (1-2)(2-3)(3-4)(4-5) L (1-2)(2-3)(3-4)(4-5) L
TOPAIS	10	13	17
*ligações repetidas ** subestruturas repetidas L - Linha R - Ramo A - Anel			

Esquema 3. Representação parcial da geração de seqüências do p-nitrotolueno.

Simultaneamente à classificação, são calculados os índices de conectividade ( ${}^n\chi$ ) e de conectividade de valência ( ${}^n\chi^v$ ) com base nas equações descritas anteriormente. Os índices de Wiener e de Schultz são calculados diretamente das matrizes distâncias e adjacências, respectivamente, sendo todos os valores acumulados numa matriz resultados que é gravada em arquivos texto (TXT) (Esquema 4) e também em arquivos tipo tabela de banco de dados (DBF).

## RESULTADOS E DISCUSSÃO

Os resultados do processamento feito com a estrutura do p-nitrotolueno são mostrados no Esquema 4, sendo que o seu gráfico molecular apresenta 93 subestruturas linha, 4 subestruturas ramo, 99 subestruturas linha-ramo e 10 subestruturas anel. O tempo de processamento para esta molécula é de aproximadamente 3 segundos quando calculada em um microcomputador pentium de 150 MHz. Um exemplo de arquivo de saída do programa pode ser visto no Esquema 4.

A medida que o número de átomos na cadeia cresce, o tempo de processamento cresce exponencialmente e isso pode ser verificado com um alcalóide do ergot (Figura 5).

Arquivo: NITROTOL.TXT  
Fórmula Molecular = C<sub>7</sub>H<sub>7</sub>NO<sub>2</sub>  
Comentários: p-nitrotolueno  
Tarefa:

10 átomos pesados na molécula  
10 ligações entre átomos pesados  
13 subestruturas com 2 ligações  
17 subestruturas com 3 ligações  
22 subestruturas com 4 ligações  
29 subestruturas com 5 ligações  
35 subestruturas com 6 ligações  
35 subestruturas com 7 ligações  
25 subestruturas com 8 ligações  
9 subestruturas com 9 ligações  
1 subestruturas com 10 ligações

### Índices de conectividade

LINHA		Chi	Chi-Val(V)
Ord	N		
0	10	7,5605	5,5731
1	10	4,6984	2,9101
2	13	4,2640	2,0927
3	14	3,0033	1,2448
4	14	1,9525	0,6665
5	16	1,4491	0,4413
6	12	0,9200	0,1764
7	4	0,1925	0,0203

RAMO		Chi	Chi-Val(V)
Ord	N		
3	3	0,7887	0,2785
5	1	0,1667	0,0124

LINHA/RAMO		Chi	Chi-Val(V)
Ord	N		
4	8	1,4487	0,3824
5	12	1,4042	0,3148
6	22	1,6649	0,3291
7	29	1,7659	0,2163
8	22	1,1382	0,0876
9	6	0,2887	0,0124

ANEL		Chi	Chi-Val(V)
Ord	N		
6	1	0,0833	0,0278
7	2	0,1314	0,0402
8	3	0,1443	0,0226
9	3	0,1443	0,0122
10	1	0,0481	0,0021

Índice de Wiener = 134  
Índice de Schultz (MTI) = 554

Tempo Gasto: 00:00:03 h  
..... Fim do arquivo .....

Esquema 4. Arquivo de saída do programa Top do p-nitrotolueno.

O gráfico molecular correspondente possui 26 vértices com 29 lados, gerando 935.674 subestruturas com ordem de 1 a 29. Para este sistema o tempo de cálculo, no mesmo computador,

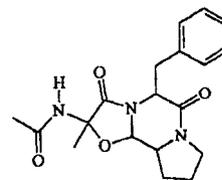


Figura 5. Alcalóide do ergot.

é de aproximadamente 19 horas.

A metodologia adotada para a geração de todas as subestruturas mostrou-se eficiente para qualquer tipo de gráfico molecular (spiro, anéis condensados e isolados, cubanos, etc.). A depuração dos resultados foi feita passo a passo, verificando-se, manualmente, as possibilidades de combinações de moléculas menores e, no caso da estrutura de um alcalóide do ergot, verificou-se coincidência dos dados com os disponíveis da literatura<sup>23</sup> até a ordem 6, sendo o número de combinações das demais ordens aceito como correto por extensão.

Observou-se, também, que os valores numéricos obtidos para os índices de conectividade de maior ordem, podem apresentar uma certa dispersão numérica devido ao valor individual de cada seqüência ser muito pequeno, de tal modo que, mesmo ocorrendo um grande número de combinações, numa dada ordem, o somatório é menor do que 10<sup>4</sup>.

## AGRADECIMENTOS

Ao CNPq pelo auxílio financeiro e concessão de bolsa de iniciação científica e à FAPERJ pela concessão de bolsa de iniciação científica.

## REFERÊNCIAS

1. Encontra-se à disposição de pesquisadores em QSAR uma versão do arquivo executável que poderá ser transferido via e-mail. (porfirio@ufrj.br ou dacosta@ufrj.br)
2. Martin, Y. C.; *J. Med. Chem.* **1981**, *24*, 229.
3. B. Testa, L. B. Kier; in *Reviews in Computational Chemistry*, K. B. Lipkowitz, D. B. Boyd, Ed., VCH, New York **1991**.
5. L. B. Kier, L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York **1976**.
6. Kier, L. B., Hall, L. H.; *J. Pharm. Sci.* **1976**, *65*, 1806.
7. Kier, L. B., Hall, L. H.; Murray, W. J.; *J. Pharm. Sci.* **1975**, *64*, 1974.
8. Kier, L. B., Hall, L. H., Murray, W. J.; *J. Pharm. Sci.* **1976**, *65*, 1226.
9. Murray, W. J., Kier, L. B., Hall, L. H.; *J. Pharm. Sci.* **1975**, *64*, 1978.
10. Kier, L. B., Hall, L. H.; *J. Pharm. Sci.* **1981**, *70*, 930.
11. Sabljic, A.; *J. Chromatogr.* **1985**, 319.
12. Randic, M.; *J. Chromatogr.* **1978**, *1*, 161.
13. Kier, L. B., Hall, L. H., Murray, W. J.; *J. Chromatogr.* **1975**, *64*, 1971.
14. Di Paolo, T.; *J. Pharm. Sci.* **1978**, *67*, 564.
15. Di Paolo, T.; *J. Pharm. Sci.* **1979**, *68*, 39.
16. Kier, L. B., Hall, L. H.; *J. Med. Chem.* **1977**, *20*, 1631.
17. Richard, A. J., Kier, L. B.; *J. Pharm. Sci.* **1981**, *69*, 124.
18. Kier, L. B.; *J. Pharm. Sci.* **1980**, *69*, 416.
19. Domenech, R. G., March, F. J. G., Soler, R. M.; *Quant. Struct.-Act. Relat.* **1996**, *15*, 201.

20. Dang, P., Madan, A. K.; *J. Chem. Inf. Comput., Sci.* **1994**, *34*, 1162.
21. Seybold, P. G., May, M., Bagal, U. A.; *J. Chem. Educ.* **1987**, *64*, 575
22. Sabljic, A., Horvatic, D.; *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 292.
23. Takeuchi, K., Kuroda, C., Ishida, M.; *J. Comput. Chem.* **1989**, *10*, 380.
24. Bersuker, I. B., Dimoglo, K. B., Lipkowitz, D. B. Boyd, A. S.; in *Reviews in Computational Chemistry*, Ed. VCH, N.Y. 1991.
25. a) Wiener, H.; *J. Am. Chem. Soc.* **1947**, *69*, 17. b) Wiener, H.; *J. Am. Chem. Soc.* **1947**, *69*, 2636. c) Wiener, H.; *J. Chem Phys.* **1947**, *15*, 766. d) Wiener, H.; *J. Chem Phys.* **1948**, *52*, 1082.
26. Platt, J. R.; *J. Phys. Chem.* **1952**, *56*, 328.
27. a) Randic, M.; *J. Am. Chem. Soc.* **1975**, *97*, 6609. b) Randic, M., Wilkins, C. L.; *Chem. Phys. Lett.* **1979**, *63*, 332.
28. Kier, L. B., Hall, L. H.; *Molecular Connectivity in Structure-Activity Analysis*, Reserch Studies Press, Letchworth, England, 1986.
29. Kier, L. B., Hall, L. H.; *J. Pharm. Sci.* **1983**, *72*, 1170.
30. a) Schultz, H. P.; *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227. b) Nikolic, S., Trinajstic, N., Mihalic, Z.; *J. Mathem. Chem.* **1993**, *12*, 252.
31. a) Dewar, M. J. S., Thiel, W.; *J. Am. Chem. Soc.* **1977**, *15*, 4899. b) Clark, T.; *A Handbook of Computational Chemistry. A pratical Guide to Chemical Structure and Energy Calculation*, Willey-Interscience, N.Y. 1985.
32. PCModel versão 4.0 - Serena Software.
33. Formato de arquivo gerado pelo PCModel versão 4.0.