communications in mathematical
match
and in computer chemistry

no. 49, September 2003
MATCDY (49) 7- 14 (2003)

ISSN 0340 - 6253

Lipophilic connectivity indices

Črtomir Podlipnik [a,*], Tomaž Šolmajer [b] and Jože Koller [a]

[a] Faculty of Chemistry and Chemical Technology, Aškerčeva 5, 1000 Ljubljana, Slovenia

[b] National Institute of Chemistry, Hajdrihova 35, 1000 Ljubljana, Slovenia

Recieved January 16, 2003

Abstract

In this paper the methodology for computation of a new type of lipophilic connectivity indices is presented. The new indices contain information about topology and molecular lipophilicity of molecular systems.

1 Introduction

Molecular lipophilicity is an important molecular property that is usually represented as a logarithm of partition coefficient in n-octanol/water system (logP). The logP is a physicochemical descriptor that is used in many quantitative structure-activity relationships (QSAR) studies for modeling biochemical and pharmacokinetic processes, transports across biological membranes, and toxicity of organic compounds [1]. The studies of Audry [2], Fauchere [3] and Furet [4] have clearly demonstrated the ability of the molecular lipophilicity potential (MLP) to describe qualitatively the 3D distribution of lipophilicity, either in space or on a molecular surface. This 3D description of molecular lipophilic potential is very valuable in 3D QSAR methodologies. In this study we tried to develop new descriptors that contain information about topology and molecular lipophilicity of molecular systems. The basic idea rises from the theory of quantum topological similarity that was introduced by Carbo et al. [5]. We restrict ourselves to the

*To whom correspodence should be addressed. Email address: crtomir.podlipnik@uni-lj.si (Črtomir Podlipnik).

Randić type molecular indices [6]. However, structure of the MLP matrices leaves open the possibility to calculate also some other types of indices like: Wiener, Shultz, Harary, Balaban, etc. [7]. The molecular lipophilicity potential (MLP) was first introduced by Audry et al. [2]. They postulated MLP (in analogy to the electrostatic potential function) in a functional form:

$$\mathrm{MLP}(\mathbf{r}) = \sum_{i=1}^{N} f_i F(|\mathbf{r} - \mathbf{R}_i|), \tag{1}$$

where f_i is partial lipophilicity of the i-th fragment and $F(|\mathbf{r} - \mathbf{R}_i|)$ is some distance function between the measured point in space and the i-th fragment.

A variety of fragmentation systems [8, 9, 10, 11] and different distance functions [2, 3, 12] have been proposed in the definition of MLP.

In our study we used MLP based on atomic lipophilic system of Broto et al. [11] and distance function $\exp(-|\mathbf{r} - \mathbf{R}_i|)/2)$ proposed by Gaillard et al. [12]. The distance function was then approximated by two Gaussian functions due to faster computation of MLP overlap integrals.

$$\exp(-r/2) \approx 0.4120 \exp(-1.2843r^2) + 0.5091 \exp(-0.0848r^2) \tag{2}$$

The scheme proposed by Broto et al. [11] was used for decomposition of molecule to atomic fragments and assignment of partial lipophilicity of this fragments. This decomposition scheme is based on decomposition of 1868 logP values into 222 atomic contributions in order to take into account nature of atoms and the connected bond types. The only information that is necessary for calculation of atomic contributions of logP is the connectivity table. Molecular geometries that are needed for MLP calculation have been obtained using molecular structure generator - CORINA [13].

2 Definition of the lipophilic connectivity indices

A graph-theoretical approach to QSAR is based on using the topological indices for encoding the structural information. In classical topological approach vertex-adjacency matrix **A** is defined in such way that matrix element A_{ij} has a value 1 or 0 if atoms i and j are bonded or not, respectively. The valences v_i of the i-th vertex in graph are then defined as the sum of i-th row elements in matrix **A**. The Randić index can be written as:

$$\chi = \sum_{bonds} \frac{A_{ij}}{\sqrt{v_i v_j}}. \tag{3}$$

Carbo et al. [5] have introduced concepts of topological quantum similarity indices. In these studies the adjacency matrix was replaced by overlap matrix **T**, in particular case, elements

of matrix are overlaps between 1 GTO (Gaussian type orbital) functions that describe atomic densities. In our study, the classical topological matrix was replaced with MLP overlap matrix. Three different types of elements occur in this matrix: T_{ij}^{++} and T_{ij}^{--} which represent overlap between vertices i and j that both have the same sign of partial lipophilicity; T_{ij}^{+-} where overlap between vertices i and j have opposite sign of partial lipophilicity. Negative values T_{ij}^{+-} in MLP overlap matrix are the reason that valence of vertex may obtain negative value. To overcome this problem the following approach was introduced. **T** matrix was decomposed into components: $\mathbf{T}^{++}$, $\mathbf{T}^{--}$ and $\mathbf{T}^{+-}$:

$$\mathbf{T} = \mathbf{T}^{++} + \mathbf{T}^{--} - \mathbf{T}^{+-}. \tag{4}$$

The valencies v_i^{++}, v_i^{--} and v_i^{+-} of i-th atom are then caluclated by sum of i-th row element in matrices $\mathbf{T}^{++}$,$\mathbf{T}^{--}$ and $\mathbf{T}^{+-}$, respectively.

The three new descriptors were obtained χ^{++}, χ^{--} and χ^{+-}:

$$\chi^{++} = \sum_{i<j} \frac{T_{ij}^{++}}{\sqrt{v_i^{++} v_j^{++}}}, \qquad v_i^{++} v_j^{++} > 0, \tag{5}$$

$$\chi^{--} = \sum_{i<j} \frac{T_{ij}^{--}}{\sqrt{v_i^{--} v_j^{--}}}, \qquad v_i^{--} v_j^{--} > 0, \tag{6}$$

$$\chi^{+-} = \sum_{i<j} \frac{T_{ij}^{+-}}{\sqrt{v_i^{+-} v_j^{+-}}}, \qquad v_i^{+-} v_j^{+-} > 0. \tag{7}$$

Descriptor χ^{++} (χ^{--}) describes intramolecular interaction between lipophilic (hydrophilic) parts of molecule while χ^{+-} describes intramolecular interactions between lipophilic and hydrophilic parts of molecule. Two other 'combined' descriptors χ_p and χ_t were introduced:

$$\chi_p = \chi^{++} + \chi^{--}. \tag{8}$$

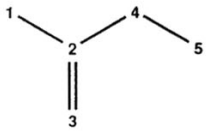
$$\chi_t = \sqrt{\chi_p^2 + (\chi^{+-})^2}. \tag{9}$$

The descriptor χ_p represents the contribution to intramolecular interactions between fragments that have the same (the opposite) sign of partial lipophilicity. The descriptor χ_t is topological selfsimilarity index that contains the information about total lipophilicity.

3 Calculation of lipophilic connectivity indices - molecular example

We will illustrate our approach on hydroxypropanone molecule. Alphanumeric code developed by Croizet et al. [14] was used for assigning atomic lipophilicity constants based on the fragment lipophilicity contributions published by Broto et al. [11]. The hydroxypropanone consists of two lipophilic fragments: methyl (C1) at the position 1 and methylene (C11) at the position 4 and three fragments that show hydrophilic characteristics: carbon (C112) at the position 2 and oxygen (O2C11) at the position 3 of carboxy group and hydroxyl (O1C1) group at the position 5. Table 1 shows the results of Broto-Moreau fragmentation of hydroxypropanone.

Table 1: The Moreau-Broto fragmentation and calculation of logP for hydroxypropanone molecule.



No.	code	f_i
1	C1	0.6310
2	C112	-0.5480
3	O2C11	-0.6810
4	C11	0.4560
5	O1C1	-0.9900
logP		-1.1320

The connectivity matrix **T** for molecule of hydroxypropanone is shown by equation 10:

$$\mathbf{T} = \begin{pmatrix} 0. & & & & \\ -7.0315 & 0. & & & \\ -7.5173 & 7.8880 & 0. & & \\ 4.7551 & -5.0814 & -5.4325 & 0. & \\ -7.3258 & 9.4058 & 11.1563 & -9.2798 & 0. \end{pmatrix} \tag{10}$$

Non-diagonal elements of T(i, j) of the matrix **T** are defined by following overlap integral:

$$\mathrm{T(i,j)} = f_i f_j \int \sum_{k=1}^{2} \sum_{l=1}^{2} a_k a_l \exp\left(-b_k \mathbf{r}_i^2\right) \exp\left(-b_l \mathbf{r}_j^2\right) d\mathbf{r}, \tag{11}$$

values a_k, a_l, b_k, b_l are taken from equation 2. The diagonal elements T(i, i) are set to zero. The overlap integrals over the Gaussians can be solved easily by using Gaussian product theorem:

$$\int \exp\left(-b_k \mathbf{r}_i^2\right) \exp\left(-b_l \mathbf{r}_j^2\right) d\mathbf{r} = \left(\frac{\pi}{b_k + b_l}\right)^{3/2} \exp\left(\frac{-b_k b_l}{b_k + b_l} |\mathbf{r}_i - \mathbf{r}_j|^2\right), \tag{12}$$

where $|\mathbf{r}_i - \mathbf{r}_j|$ is distance between i-th and j-th atom in molecule. MLP valences v are calculated by summation of rows in matrix $\mathbf{T}$:

$$\mathbf{v} = (-17.1194, 5.1808, 6.0946, -15.0386, 3.9565) \tag{13}$$

The similar methodology is used for calcualton of $\mathbf{T}^{++}$, $\mathbf{T}^{--}$,$\mathbf{T}^{++}$ and the corresponding MLP valences vectors $\mathbf{v}^{++}$,$\mathbf{v}^{--}$, $\mathbf{v}^{+-}$ of these matrices.

$$\mathbf{T}^{++} = \begin{pmatrix} 0. & & & & \\ 00000 & 0. & & & \\ 0.0000 & 0.0000 & 0. & & \\ 4.7551 & 0.0000 & 0.0000 & 0. & \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0. \end{pmatrix} \tag{14}$$

$$\mathbf{v}^{++} = (4.7551, 0.0000, 0.0000, 4.7551, 0.0000) \tag{15}$$

$$\mathbf{T}^{--} = \begin{pmatrix} 0. & & & & \\ 0.0000 & 0. & & & \\ 0.0000 & 7.8880 & 0. & & \\ 0.0000 & 0.0000 & 0.0000 & 0. & \\ 0.0000 & 9.4058 & 11.1563 & 0.0000 & 0. \end{pmatrix} \tag{16}$$

$$\mathbf{v}^{--} = (0.0000, 17.2938, 19.0443, 0.0000, 20.5621) \tag{17}$$

$$\mathbf{T}^{+-} = \begin{pmatrix} 0. & & & & \\ 7.0315 & 0. & & & \\ 7.5173 & 0.0000 & 0. & & \\ 0.0000 & 5.0814 & 5.4324 & 0. & \\ 7.3258 & 0.0000 & 0.0000 & 9.2798 & 0. \end{pmatrix} \tag{18}$$

$$\mathbf{v}^{+-} = (21.8745, 12.1129, 12.9497, 19.7939, 16.6055) \tag{19}$$

From tabulated values in table 1, and equations 5-9, the Randić indices can be computed adding a term for each bond. The final results of these computations are collected in the table 2.

Table 2: Results of the lipophilic connectivity index computation for hydroxypropanone molecule.

Index	Value
χ^{++}	1.0000
χ^{--}	1.4972
χ^{+-}	2.4423
χ_p	2.4972
χ_c	3.4929

The new lipophilic connectivity descriptors have been calculated for the three analogous series of compounds (Fig. 1): carboxyl acids (1-4); 3-hydroxy-alkane-2-ones (5-8); 1-hydroxy-alkane-2-ones (9-12).

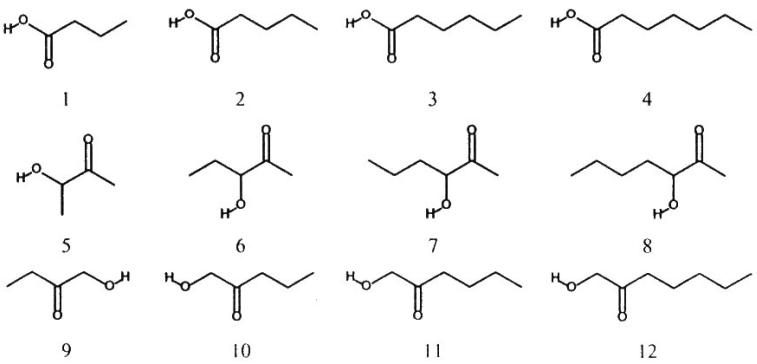


Figure 1: The structures of compounds used in our study.

Results of the calculation of different lipophilic connectivity descriptors for these compounds are shown in the table 3.

Table 3: Topological selfsimilarity indices of carboxyl acids (1-4); 3-hydroxy-alkane-2-ones (5-8); 1-hydroxy-alkane-3-ones (9-12).

No.	χ^{+-}	χ^{++}	χ^{--}	χ_p	χ_c
1	2.8196	1.9892	1.0000	2.8196	4.1092
2	3.1254	2.4819	1.0000	3.1254	4.6789
3	3.3617	2.9738	1.0000	3.3617	5.2050
4	3.5384	3.4657	1.0000	3.5384	5.6976
5	2.6942	1.2848	1.4982	2.6942	3.8735
6	3.1967	1.7582	1.4982	3.1967	4.5632
7	3.6102	2.2361	1.4982	3.6102	5.1941
8	3.9355	2.7191	1.4982	3.9355	5.7684
9	2.9950	1.4894	1.4972	2.9950	4.2296
10	3.4429	1.9779	1.4972	3.4429	4.8919
11	3.8011	2.4673	1.4972	3.8011	5.4924
12	4.0691	2.9570	1.4972	4.0691	6.0331

Since all compounds in the same analogous series have a similar spatial distribution of hydrophilic groups this results in almost the same values of the descriptor χ^{--}. The general trends of all other descriptors is that they increase with length of alkyl chain (the lipophilic part of molecules).

4 Conclusions

The method for calculating molecular descriptors, which combine atom-atom connectivity with a molecular lipophilic potential is presented. We introduced lipohilic connectivity descriptors χ^{+-}, χ^{++} and χ^{--} that describe various types of lipophilic intramolecular interactions. The next two 'combined' descriptors χ_p and χ_c were composed from previous three descriptors. The method for calculating descriptors is fast. The only information that we need for the calculation of these indices is atomic logP increments, molecular geometry and the corresponding connectivity table.

References

[1] A. Carpy, *Analusis* 27 (1999) 3-6.

[2] E. Audry, J.P. Dubost, J.C. Colleter and P. Dallet, *Bur. J. Med. Chem.* 21 (1986) 71-72.

[3] J.L. Fauchere, P. Quarendon, P. and L. Kaetterer, *J. Mol. Graphics* 6 (1988) 202-206.

[4] P. Furet, A. Sele and N.C. Cohen, *J. Mol. Graphics* 6 (1988) 182-189.

[5] E. Besalu, A. Gallegos and R, Carbo-Dorca, *MATCH Commun. Math. Comput. Chem.* 44 (2001) 41-64.

[6] M. Randić, *J. Am. Chem. Soc* 97 (1975) 6609-6615.

[7] N. Trinajstić, Chemical Graph Theory, 2nd revised edition, (CRC Press, Boca Raton, Florida, 1992)

[8] R.F. Rekker and H.M. De Kort, *Eur. J. Med. Chem.* 14 (1979) 479-488.

[9] C. Hansch and A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, (Wiley, New York, NY , 1979)

[10] A.K. Ghose and G.M. Crippen, *J. Comput. Chem.* 7 (1986) 565-577.

[11] P. Broto, G. Moreau and C. Vandycke, *Eur. J. Med. Chem.* 19 (1984) 61-65.

[12] P. Gaillard, P. Carrupt, B. Testa and A. Boudon, *J. Comput.-Aid. Mol. Design* 8 (1994) 83-96.

[13] J. Gasteiger, C. Rudolph and J. Sadowski, *Tetrahedron Comp. Method* 3 (1990) 537-547.

[14] F. Croizet, J.P. Dubost, M.H. Langlois and E. Audrey, *Quant. Struct.-Activ. Relat.* 10 (1991) 211-215.