Buletinul Ştiinţific al Universităţii "Politehnica" din Timisoara, ROMÂNIA
Seria CHIMIE ŞI INGINERIA MEDIULUI
Chem. Bull. "POLITEHNICA" Univ. (Timişoara) *Volume 51(65), 1-2, 2006*

New Shape Descriptors for Quantitative Treatment of Steric Effects. III. A New Globularity Measure for QSPR/QSAR Studies

Mihai Medeleanu*, Dan Ciubotariu**, Ciprian Ciubotariu***

* POLITEHNICA University of Timişoara, Applied Chemistry and Organic and Natural Compounds Engineering Department, Timişoara, C. Telbisz 6, Romania, e-mail: mihai.medeleanu@chim.upt.ro

**University of Medicine and Pharmacy, Organic Chemistry Department, Timişoara, P-ta. E. Murgu 2, Romania

*** POLITEHNICA University of Timişoara, Computers Department, Timişoara, Bd. Vasile Pârvan 2, Romania

Abstract: Size and shape of the organic molecules are useful descriptors in QSPR/QSAR studies. Usual characterization of molecular size is based on steric parameters like van der Waals, STERIMOL, Charton or Taft types. The main problem with the second molecular property is the quantification of shape. In this paper we propose a new shape descriptor based on van der Waals parameters like volume and surface. The computation algorithm and a set of molecular examples are depicted.

1. Introduction

The modern studies concerning the quantitative structure-properties relationships (QSPR) and those concerning the quantitative structure-activity relationships (QSAR) are focused on the quantification of both the molecular structure and the physical, chemical properties or biological activity. This procedure requires a separation of the effects involved in chemical or biological interactions, an independent quantification of these as electronic, steric, hydrophobic, or other types of factors and the quantitative treatment of reactivity or biological activity by means of the correlation analysis models. The use of the extrathermodynamic formalism made possible the theoretical validation of this approach [1].

It is known that van der Waals parameters are useful for modeling the steric effects of organic molecules. The van der Waals volume and surface are largely used in studies like packing density of molecules[2], protein folding[3], quantification of biological activity[4-6]. Another derived parameter was presented in an early stage of this study and defined as "molecular globularity"[7], a useful shape descriptor of molecular structure. In this paper, a new and very simple globularity parameter will be defined together with some applications and a comparison with the former one.

2. Computational Methods and Results

Various definitions of the globularity can be found in the literature, mainly starting from two basic van der Waals descriptors: the van de Waals volume and the van der Waals surface. The volume can be computed by analytical methods[8] or by Monte Carlo algorithms[9,10]. In this way a molecular van der Waals envelope, Γ, can be defined as the external surface resulted following the interaction of all vdW spheres, corresponding to the atoms of the molecule M. The points (x,y,z) inside the envelope satisfy at least one of the following inequalities:

$$(X_i - x)^2 + (Y_i - y)^2 + (Z_i - z)^2 \le (r_i^w)^2 \quad i = \overline{1,m} \tag{1}$$

where m represents the number of atoms of M. Consequently the total volume from inside the envelope represents the molecular vdW volume (V_M^w) of the molecule M.

Let be a function F=f(x,y,z), continuous in a bordered and closed domain M. The integral:

$$I = \iiint_M f(x,y,z) \times dx \times dy \times dz \tag{2}$$

can be interpreted as a weight, f(x,y,z) being the repartition of density in space M. If we consider the function f(x,y,z) (the density) is identical equal with 1, than the following integral:

$$V^w = \iiint_M dV \qquad dV = dx \times dy \times dz \tag{3}$$

can be intuitively justified as a volume. This assumption is natural because the properties of molecular vdW space can be considered independent from the nature of the atoms, even in the case of intersection domains between the vdW atomic spheres.

One of the main problem is to find out the volume of the intersection of the spheres in order to avoid multiple computations of the same volume.

To estimate the integral (3), the spheres that models the molecule are inserted into a parallelepiped with the volume V_p. The random numbers are generated into the domain M, which include the parallelepiped. If n_t is the total number of generated points and n_s the number of points that satisfies the equations (1), than the van der Waals volume is:

$$V_M^w = \frac{n_s}{n_t} \times V_p \tag{4}$$

A similar method is used for the van der Waals surface. The algorithm implies the generation of a uniform grid[11] on each sphere of the molecule, followed by the detection from the number of all points generated on the surface (n_t) of those (n_e) that satisfies the *equality* in equations (1). For every "hard sphere" i, one can compute the relation:

$$S_i^w = \frac{(n_e)_i}{n_t} 4\pi (r_i^w)^2 \tag{5}$$

The final surface is computed after the elimination of overlays (S_k):

$$S^w = \sum_{i=1}^{m} S_i^w - \sum_{k} S_k \tag{6}$$

The full algorithms and results are described in some previous papers on this subject [9-11].

The molecular globularity is defined as the ratio between the surface of an equivalent sphere (the sphere with the volume equal to molecular volume) and the molecular surface [12], or the inverse of the above ratio [13]. A value of 1.0 means a shape of a perfect sphere.

With the van der Waals volume and surface computed, two other parameters can be defined. The first is the ratio between the volume and the surface of the molecule, which has the dimensions of a length:

$$L^w = \frac{V^w}{S^w} \tag{7}$$

The second one, defined only for acyclic molecules is named the globularity measure (G_m) and is given by the relation:

$$G_m = \frac{L^w}{R_s} \tag{8}$$

where L^w is defined by relation (7) and R_s represents the ratio between the volume and the surface of an equivalent sphere which surround the molecule, with the radius equal to the half of the longest dimension of the parallelepiped which circumscribes the molecule. For cyclic molecules, the above relation cannot be used because the volume of the equivalent sphere includes the internal empty space (for instance a cavity), which is not included in the van der Waals volume. The results for the globularity of such molecules are far from reality (for C60 fullerene G_m=0.686).

These two parameters can be used to describe the shape of the acyclic molecules. The globularity measure decreases with the growth of the linear chains and increases toward unity when the molecule is highly branched or compacted. In Table 1 the van der Waals volumes, surfaces, L^w and G_m parameters are listed for C2-C8 alkanes (M represents the methyl group, and the numbers before reflect th position of branching).

TABLE 1. The van der Waals parameters for C2-C8 alkanes

Alcane	**V^w**	**S^w**	**L^w**	**G_m**
C2	44.205	70.920	0.623	0.653
C3	60.234	93.026	0.647	0.554
C4	77.845	115.315	0.675	0.516
2M-C3	77.275	114.675	0.674	0.558
C5	94.749	138.127	0.686	0.451
2M-C4	94.047	136.241	0.690	0.522
22MM-C3	92.368	135.225	0.683	0.527
C6	110.242	159.338	0.692	0.417
2M-C5	109.970	158.477	0.694	0.465
3M-C5	109.676	156.114	0.703	0.453
22MM-C4	110.161	155.015	0.711	0.537
23MM-C4	110.215	155.033	0.711	0.496
C7	127.399	181.834	0.701	0.375
2M-C6	127.317	179.742	0.708	0.429
3M-C6	126.956	178.780	0.710	0.448
3E-C5	125.874	176.827	0.712	0.458
22MM-C5	126.822	176.663	0.718	0.486
33MM-C5	125.592	174.301	0.721	0.451
24MM-C5	127.320	178.291	0.714	0.462
23MM-C5	127.492	174.419	0.731	0.472
223MMM-C4	126.616	172.912	0.732	0.511
C8	142.613	203.964	0.699	0.347
2M-C7	143.144	201.813	0.709	0.388
3M-C7	143.774	200.463	0.717	0.373
4M-C7	143.925	199.927	0.720	0.401
3E-C6	141.739	198.490	0.714	0.422
22MM-C6	142.733	198.900	0.718	0.443
33MM-C6	142.566	195.458	0.729	0.410
23MM-C6	142.990	195.654	0.731	0.429
24MM-C6	144.186	197.986	0.728	0.429
25MM-C6	143.512	198.892	0.722	0.396
34MM-C6	142.278	195.908	0.726	0.456
234MMM-C5	142.593	196.954	0.724	0.465
223MMM-C5	142.505	193.130	0.738	0.482
224MMM-C5	142.473	197.110	0.723	0.478
233MMM-C5	141.973	191.844	0.740	0.468
3E-2M-C5	142.069	195.215	0.728	0.481
3E-3M-C5	142.811	192.962	0.740	0.473
2233MMMM-C4	142.466	189.633	0.751	0.522

As expected, the longer the chain, the smaller the globularity measure. This measure increases for alkanes such as 2,2- and 2,3-dimethylbutane, 2,2,3-trimethylbutane, trimethylpentanes, ethyl-methyl-pentanes or tetrametrhylbutanes.

An useful example which describes the utility of the globularity measure is in the case of molecules with long chain, which can have different conformations. For instance, let be the normal C16 alkane, with two conformations presented below (figures 1 and 2).

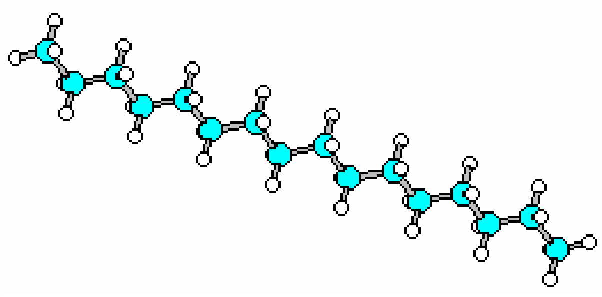

Fig. 1. The linear conformation of C16 alkane

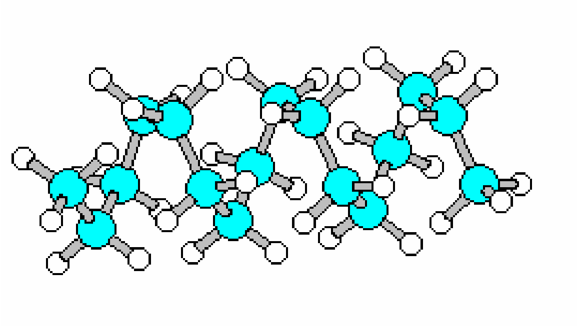

Fig. 2. A twisted conformation of C16 alkane

TABLE 2. van der Waals parameters for the two conformations of C16 alkane

Alcane	V^w	S^w	L^w	G_m
C16 normal	283.63	381.51	0.743	0.214
C16 twisted	244.04	249.15	0.980	0.448

The described parameters could be used in the globularity studies of macromolecules like polypeptides, proteines and others macromolecules.

In order to avoid the problems raised by the described globularity measure (G_m) and to include the cavities volume for molecules with this kind of structure, another molecular descriptor has been defined: the ellipsoidal globularity G_e.

For the computation of this molecular descriptor the molecule is surrounded by an ellipsoid tangent to the extremities. The ellipsoidal shape is a better approach than the parallelepiped use in former computations. Considering the longest half axis of the ellipsoid, the corresponding sphere can be setup. The ratio between the volume of the ellipsoid and the volume of the sphere defines the ellipsoidal globularity:

$$G_{el} = \frac{V^{el}}{V^s} \tag{9}$$

Knowing the volume of a sphere:

$$V^s = \frac{4\pi}{3} \cdot R^3 \tag{10}$$

and that of an ellipsoid:

$$V^{el} = \frac{4\pi}{3} abc \tag{11}$$

where R is the sphere radius, a,b,c the half of the ellipsoidal axes, and

$$R = \max(a, b, c) \tag{12}$$

It is obvious that the ellipsoidal globularity is much easier to compute. The van der Waals volume is not necessary.

TABLE 3. The globularity measures

molecule	V^S	V^{el}	G_m	G_{el}
C2	130.30	114.65	0.613	0.880
C3	226.23	160.41	0.534	0.709
C4	310.33	190.72	0.494	0.615
C5	469.06	247.29	0.438	0.527
C6	603.19	284.80	0.408	0.472
C7	843.12	352.73	0.368	0.418
C8	1038.19	397.35	0.346	0.383
C9	1376.35	476.58	0.317	0.346
C10	1644.28	528.17	0.300	0.321
C15	4215.23	958.81	0.223	0.227
C20-fullerene	251.93	222.63	0.805	0.884
22MM-C3	303.30	254.81	0.513	0.840
2M-C3	246.28	217.79	0.536	0.884
22MM-C4	318.27	311.76	0.518	0.980
23MM-C4	382.21	332.28	0.488	0.869
2M-C4	321.10	264.91	0.504	0.825
22MM-C5	462.81	390.69	0.461	0.844
23MM-C5	480.68	397.14	0.461	0.826
24MM-C5	532.12	363.28	0.438	0.683
2M-C5	466.43	339.18	0.449	0.727
33MM-C5	499.17	373.21	0.457	0.748
3E-C5	502.22	390.36	0.450	0.777
3M-C5	492.65	318.16	0.446	0.646
2M-C6	610.12	389.06	0.414	0.638
3M-C6	641.43	369.41	0.411	0.576
C60	681.62	665.85	0.686	0.977
Anthracene	1197.43	293.22	0.382	0.245
Benzene	371.61	152.67	0.512	0.411
Naphthalene	709.72	218.38	0.440	0.308
Phenanthrene	1217.99	261.84	0.384	0.215
Pyrene	1195.52	292.85	0.396	0.245

In Table 3 are listed the spherical and ellipsoidal volumes together with the corresponding globularities. As it can be seen the ellipsoidal globularity gives a much better shape description for molecules such as fullerenes, but also for highly branched alkanes. From the same table one can observe that the aromatic molecules have a lower globularity, in agreement with their real planar structure.

3. Conclusions

Two new methods to setup molecular descriptors for the molecular shape have been proposed. Their results are the globularity measure G_m and the ellipsoidal globularity G_{el}. The former is useful in shape studies for acyclic molecules (including macromolecules) and the later, a more general descriptor can be used in the shape description of any kind of molecules, including ones with cavities

Acknowledgments

The authors acknowledge Prof. Dr. Mircea Mracec from the Chemical Institute of the Romanian Academy in Timişoara for providing the HyperChem[14] package for molecular optimizations of the studied structures.

References

1. Ciubotariu, D.; Deretey, E.; Medeleanu, M.; Gogonea, V.; Iorga, I. Chem. Bull. PIT, **1990**, 35, (49), 83
2. Kitaigorodski, A.I. “Molecular Crystals and Molecules”, Academic Press, New York, **1973**
3. Lee, B.; Richards, F.M. J.Mol.Biol., **1971**, 55, 379; **1974**, 82, 1
4. Moriguchi, I.; Kanada, Y.; Komatsu, K. Chem.Pharm.Bull., **1976**, 24, 1799
5. Tichy, M. Int.J.Quantum Chem., **1979**, 16, 509
6. Palmer, R.A.; Tickle, J.H.; Tickle, J.L. J.Mol.Graphics, **1983**, 1, 94
7. Medeleanu, M. “Structure-Properties Correlations by Topological Methods”, PhD. Thesis, Univ. “Politehnica”, Bucureşti, **1997**
8. Gogonea, V.; Osawa, E. J.Comput.Chem., **1995**, 16, 817
9. Mureşan, S.; Bologa, C.; Medeleanu, M.; Gogonea, V.; Dragoş, D.; Ciubotariu, D.; Chem.Bull.Tech. Univ. (Timişoara), **1994**, 39, 47
10. Ciubotariu,D.; Medeleanu,M.; Iorga,I.; Deretey,E.; Bologa,C.; Mureşan,S., Chem.Bull.Tech.Univ.(Timişoara), **1993**, 38, 83
11. Gogonea, V.; Ciubotariu, D.; Deretey, E.; Popescu, M.; Iorga, I.; Medeleanu, M. Rev. Roum. Chim., **1991**, 36, 465
12. Filliponi, E.; Cruciani, G.; Tabarrini, O.; Cecchetti, V.; Fravolini, A. J. Comput.-Aided Mol. Design **2001**, 15, 203-217.
13. Torrens, F. Int. J. Mol. Sci. **2001**, 2, 72-88
14. Hypercube, Inc. HyperChem®Computational Chemistry, **1996**