J. Chil. Chem. Soc., 61, Nº 4 (2016)

A BOX-MODEL APPROACH TO MOLECULAR SHAPE

ADELIO R. MATAMALA

Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de Concepción.
Edmundo Larenas 129, Concepción, Chile.

ABSTRACT

In this work, using the length-width-height ratios of a rectangular box in which the molecule under study is enclosed, two descriptors for the characterization of molecular shape are introduced. The molecular shape analysis of the set of molecules under study is facilitated by the introduction of a diagram where the extreme cases (i.e. cubic-box, prolate-box and oblate-box) are clearly located. The method is illustrated by the analysis of fifty molecules including normal alkanes, acenes, alkynes, oblate molecules and the eighteen octane isomers. The geometry of each molecule was optimized by AM1 semiempirical method, and van der Waals molecular surface was used to define the molecular shape contour. The length, the width, and the height of the enclosing box were extracted using VEGA ZZ software. A regular behaviour of the shape descriptors for homologous series was observed. Finally, using the present descriptors, a molecular shape similarity analysis of the octane isomer molecules was included.

Keywords: Molecular Shape, van der Waals Molecular Surface, Molecular Similarity.

1. INTRODUCTION

Certainly, molecular shape plays a central role in Chemistry [1]. Throughout the history of Chemistry, three-dimensional representations of molecules have guided chemical intuition and rational understanding of properties and reactions. Basically, the relative arrangements of nuclei in molecules are often represented by three-dimensional structures showing the chemical bonds as lines segments interconnecting the nuclei, i.e. the so-called molecular skeleton or simply the molecular geometry. However, molecules are three-dimensional objects occupying a portion of space [2-4], i.e. molecules have steric behaviour.

According to quantum mechanics, a molecule does not have a precisely defined surface separating the system from the rest of the universe, but formal surfaces may be introduced in an approximate sense in order to define the so-called molecular surfaces. There are several ways to define a molecular surface [5]: contour surfaces of electronic density, molecular orbitals, electrostatic potentials, molecular surfaces generated by fused atomic spheres, solvent accessible surfaces, and so on; but the simplest is based on the description of the molecule as a collection of atomic fused spheres of some appropriate defined van der Waals (vdW) radii [6-8]. Positions of these spheres are described by their Cartesian coordinates according to the three-dimensional stereo-chemical bond pattern of the molecule. The envelope of the outer surface of the vdW atomic fused spheres defines the three-dimensional van der Waals (3D-vdW) representation of the molecule [9]. This simple molecular representation has been very useful for the study of gases, solids and liquids [10], and particularly for the description of ligand molecules interacting with receptors in biological systems [11]. In the present work, 3D-vdW molecular surfaces are adopted to define the molecular shape.

There are several shape descriptors for the characterization of 3D molecular shape [12] and most of them can be grouped into any of the following classifications: (i) derived from inertia tensor (inertial shape factor [13], molecular eccentricity [14], molecular asphericity [14], Amoore shape indices [15], etc.), (ii) derived from geometry matrix (geometrical shape coefficient [16], sphericity [17], characteristic ratio [14], etc.), (iii) derived from surface to volume ratio (ovality index [18], globularity factor [2], surface-volume ratio [19], etc.) and (iv) derived from dimension ratios (Kaliszan shape parameter [20], length-to-breadth ratio [21], sterimol parameters [22], etc.). On the one hand, the first two groups of descriptors are calculated from the geometry of molecular skeleton and provide very important information about the space arrangement of nuclei in the molecule, but unfortunately that kind of descriptors do not take into account the space extension of the electron cloud. On the other hand, descriptors of the last two categories are less sensitive to the morphological details of molecules, but that kind of parameters include in their description the space of the electron cloud.

In this work, following the underlying idea of the length-to-breadth ratio descriptors, two indexes are introduced to model the molecular shape by using of the length-width-height ratios of a rectangular box (i.e. right rectangular prism) in which the molecule under study is enclosed. This approach is termed Box-Model to Molecular Shape (BMMS), and it is an oversimplified but effective model. One advantage of the present formulation is the simplicity in the calculation of the descriptors. Other advantage is the location of extreme cases (i.e. cubic-box, oblate-box, and prolate-box) as well-defined references for comparison purposes. On the other hand, BMMS can be applied to any molecule for which any molecular surface is defined (i.e. electron density, electrostatic potentials, fused atomic sphere models, solvent accessible surfaces, etc.). Finally, BMMS is the first stage for the development of more realistic models which start from a global description of the molecule.

2. SHAPE PARAMETERS

Let us consider the minimal rectangular box containing the molecule (see Figure 1), where A, B, and C (with $A \geq B \geq C$) are the length, the width and the height of the box, respectively. In order to characterize the molecular shape, the molecule is replaced by its minimal enclosing box and its shape is characterized in terms of the length-width-height ratios. Certainly, this is a crude approximation to molecular shape but it is the starting stage for future more realistic models.

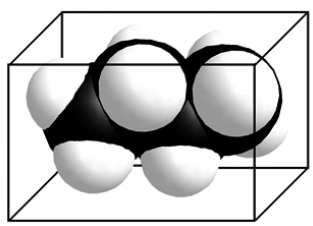

Figure 1. Molecule inside of a rectangular box.

Let us introduce the *thickness index* as the height to length ratio,

$$\tau = \frac{C}{A} \qquad (1)$$

and the *asymmetry index* as the difference between the height to width ratio and the width to length ratio of the packing box,

$$\kappa = \frac{C}{B} - \frac{B}{A} \qquad (2)$$

It is easy to see that $\tau = 1$ and $\kappa = 0$ for the cubic box, $0 < \kappa < 1$ for the rod-shape boxes with square-section, and $-1 < \kappa < 0$ for the plate-shape boxes with square-base. The limit case $\tau = 0$ and $\kappa = -1$ corresponds to the infinitely thin square plate. The limit case $\tau = 0$ and $\kappa = +1$ corresponds to the

 e-mail: adelio.matamala@udec.cl

infinitely thin rod. On the other hand, $\tau \neq 0$ and $\kappa = 0$ define boxes in which the width is the geometric mean between the length and the height, i.e. $B = \sqrt{(AC)}$. Finally, $\tau = 0$ and $\kappa = 0$ defines a singularity. Figure 2 shows graphically the above limit cases. Every pair (κ,τ) falls inside the triangle (-1,0), (1,0) and (0,1). Additionally, that triangle is divided in two parts: the prolate-box region defined by the triangle (0,0), (1,0) and (0,1), and the oblate-box region defined by the triangle (0,0), (-1,0) and (0,1).

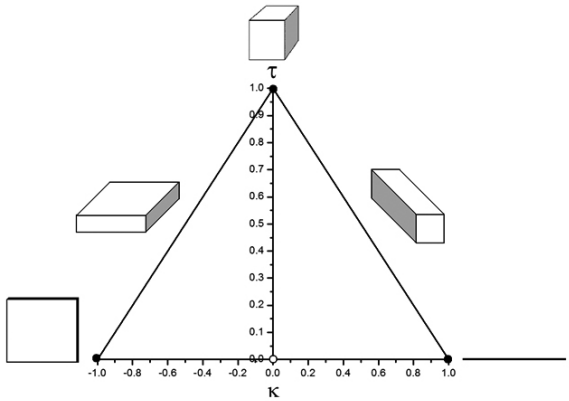

Figure 2. BMMS diagram in which the extreme cases are showed.

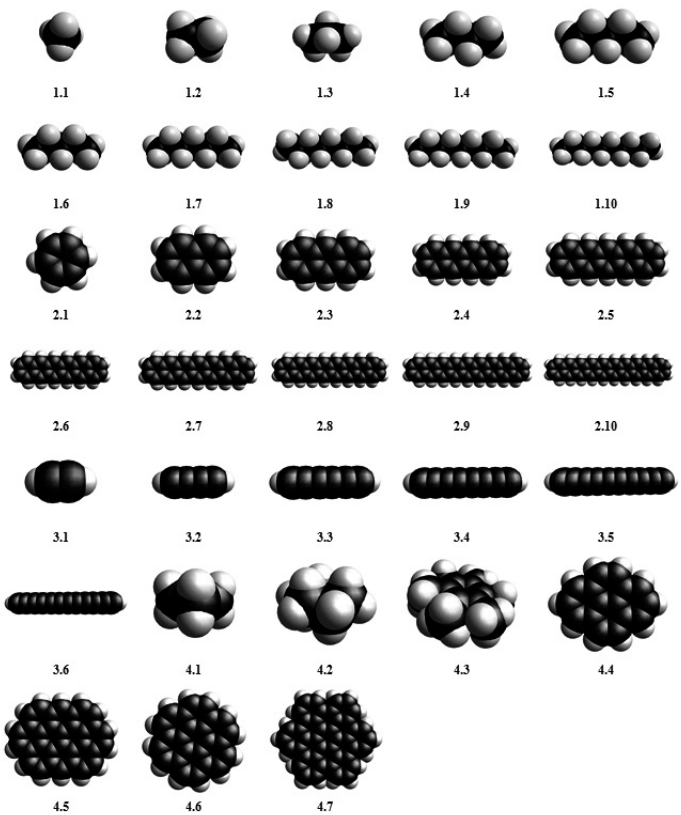

Figure 3. 3D-vdW representation of molecules from Tables 1, 2, 3 and 4. The labels are in correspondence to those appearing in the respective Tables.

3. METHOD

In order to illustrate the application of BMMS descriptors, a set of fifty molecules were analysed in terms of thickness τ and asymmetry κ shape indexes defined in (1) and (2) respectively. The geometry of each molecule was optimized using AM1 semiempirical method implemented in Gaussian 03 computational software [23]. From each optimized molecular structure, the 3D-vdW representation was calculated using VEGA ZZ software [24]. The length, the width and the height of the enclosing box were collected from molecular information output selecting 'View' and then 'Information' from the main menu.

The set of molecules was constrained to hydrocarbons in order to reduce the associated computational time, but the present method is quite general to be applied to any molecule and any level of theory. The set of fifty molecules includes: the first ten normal alkanes (see Table 1), the first ten normal acenes (see Table 2), the first six normal alkynes (see Table 3), six molecules to populate the oblate region (see Table 4), and the eighteen octane isomers (see Table 5). The first three subsets of molecules were chosen to study the behaviour of the shape parameters within homologous series. On the other hand, the collection of octane isomers represents a simple but challenging example of molecular sample with branching diversity. Figures 3 and 4 show the 3D-vdW representation for all molecules included in the present study.

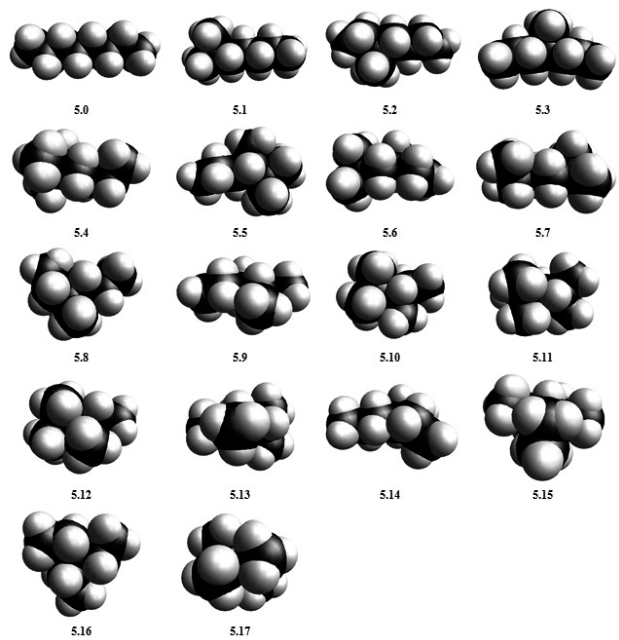

Figure 4. 3D-vdW representation of the octane isomers. The labels are in correspondence to those appearing in the Table 5.

4. RESULTS AND DISCUSSION

Tables 1, 2, 3, 4 and 5 show the length, the width and the height of each molecule, and BMMS descriptor values for each subset of molecules, respectively. Figure 5 shows the BMMS diagram for the molecules of the first four Tables. Similarly, Figure 6 shows the BMMS diagram for the octane isomers.

Table 1. Shape descriptor values for the first ten normal alkanes.

	Molecule	CAS RN	A / Å	B / Å	C / Å	κ	τ
1.1	Methane	74-82-8	4.205	4.048	4.006	0.027	0.953
1.2	Ethane	74-84-0	4.901	4.453	4.373	0.073	0.892
1.3	Propane	74-98-6	6.715	4.567	4.210	0.242	0.627
1.4	Butane	106-97-8	7.825	4.753	4.210	0.278	0.538
1.5	Pentane	109-66-0	9.214	4.587	4.210	0.420	0.457
1.6	Hexane	110-54-3	10.415	4.833	4.210	0.407	0.404
1.7	Heptane	142-82-5	11.713	4.587	4.210	0.526	0.359
1.8	Octane	111-65-9	12.938	4.819	4.210	0.501	0.325
1.9	Nonane	111-84-2	14.212	4.587	4.210	0.595	0.296
1.10	Decane	124-18-5	15.447	4.795	4.210	0.568	0.273

Table 2. Shape descriptor values for the first ten normal acenes.

	Molecule	CAS RN	A / Å	B / Å	C / Å	κ	τ
2.1	Benzene	71-43-2	7.359	6.973	3.400	-0.460	0.462
2.2	Naphthalene	91-20-3	9.178	7.404	3.400	-0.347	0.370
2.3	Anthracene	120-12-7	11.632	7.410	3.400	-0.178	0.292
2.4	Tetracene	92-24-0	14.084	7.413	3.401	-0.068	0.241
2.5	Pentacene	135-48-8	16.537	7.415	3.400	0.010	0.206
2.6	Hexacene	258-31-1	18.989	7.417	3.400	0.068	0.179
2.7	Heptacene	258-38-8	21.442	7.418	3.400	0.112	0.159
2.8	Octacene	258-33-3	23.894	7.418	3.400	0.148	0.142
2.9	Nonacene	258-36-6	26.345	7.419	3.400	0.177	0.129
2.10	Decacene	24540-30-5	28.798	7.419	3.400	0.201	0.118

Table 3. Shape descriptor values for the first six normal alkynes.

	Molecule	CAS RN	A / Å	B / Å	C / Å	κ	τ
3.1	Ethyne	74-86-2	5.717	3.400	3.400	0.405	0.595
3.2	Buta-1,3-diyne	460-12-8	8.274	3.400	3.400	0.589	0.411
3.3	Hexa-1,3,5-triyne	3161-99-7	10.826	3.400	3.400	0.686	0.314
3.4	Octa-1,3,5,7-tetrayne	6165-96-4	13.378	3.400	3.400	0.746	0.254
3.5	Deca-1,3,5,7,9-pentayne	32597-32-3	15.929	3.400	3.400	0.787	0.213
3.6	Dodeca-1,3,5,7,9,11-hexayne	32597-33-4	18.479	3.400	3.400	0.816	0.184

Table 4. Shape descriptor values for some oblate molecules.

	Molecule	CAS RN	A / Å	B / Å	C / Å	κ	τ
4.1	Cyclobutane	287-23-0	5.863	5.836	4.214	-0.281	0.719
4.2	Cyclohexane	110-82-7	7.378	6.743	5.096	-0.158	0.691
4.3	Hexamethylbenzene	87-85-4	9.147	9.091	4.282	-0.523	0.468
4.4	Pyrene	129-00-0	11.633	9.203	3.400	-0.422	0.292
4.5	Ovalene	190-26-1	14.102	11.675	3.400	-0.537	0.241
4.6	Coronene	191-07-1	11.891	11.887	3.400	-0.714	0.286
4.7	Hexabenzocoronene	190-24-9	15.939	14.163	3.400	-0.649	0.213

Table 5. Shape descriptor values for the octane isomers.

	Molecule	CAS RN	A / Å	B / Å	C / Å	κ	τ
5.0	n-octane	111-65-9	12.938	4.819	4.210	0.501	0.325
5.1	2-methyl-heptane	592-27-8	11.649	6.464	5.267	0.260	0.452
5.2	3-methyl-heptane	589-81-1	11.578	6.414	5.527	0.308	0.477
5.3	4-methyl-heptane	589-53-7	11.620	6.552	5.345	0.252	0.460
5.4	2,2-dimethyl-hexane	590-73-8	9.870	6.486	6.459	0.339	0.654
5.5	2,3-dimethyl-hexane	584-94-1	10.421	7.120	5.301	0.061	0.509
5.6	2,4-dimethyl-hexane	589-43-5	10.122	6.596	5.261	0.146	0.520
5.7	2,5-dimethyl-hexane	592-13-2	10.419	5.962	5.635	0.373	0.541
5.8	3,3-dimethyl-hexane	563-16-6	9.544	7.091	6.280	0.143	0.658
5.9	3,4-dimethyl-hexane	583-48-2	10.287	7.155	5.472	0.069	0.532
5.10	2,2,3-trimethyl-pentane	564-02-3	9.010	7.104	6.310	0.100	0.700
5.11	2,2,4-trimethyl-pentane	540-84-1	9.003	6.680	6.145	0.178	0.683
5.12	2,3,3-trimethyl-pentane	560-21-4	8.972	7.040	6.365	0.120	0.709
5.13	2,3,4-trimethyl-pentane	565-75-3	8.738	7.539	6.450	-0.007	0.738
5.14	3-ethyl-hexane	616-99-8	9.722	7.563	5.426	-0.060	0.558
5.15	3-ethyl-2-methyl-pentane	609-26-7	8.683	7.609	5.318	-0.177	0.612
5.16	3-ethyl-3-methyl-pentane	1067-08-9	8.416	8.212	6.042	-0.240	0.718
5.17	2,2,3,3-tetramethyl-Butane	594-82-1	7.244	6.722	6.667	0.064	0.920

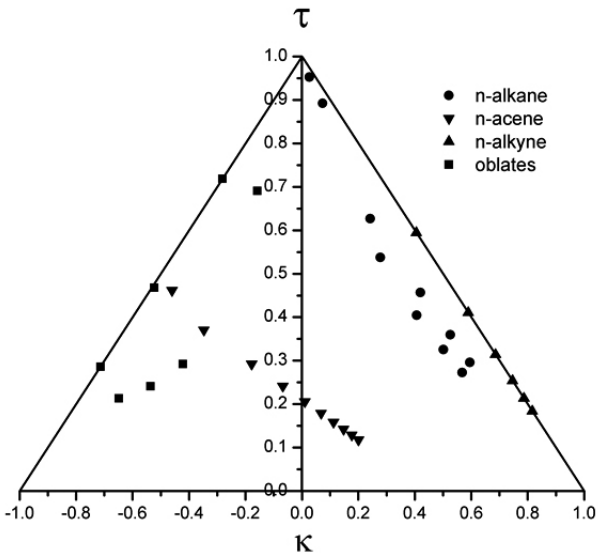

Figure 5. BMMS diagram for the molecules from Tables 1, 2, 3 and 4.

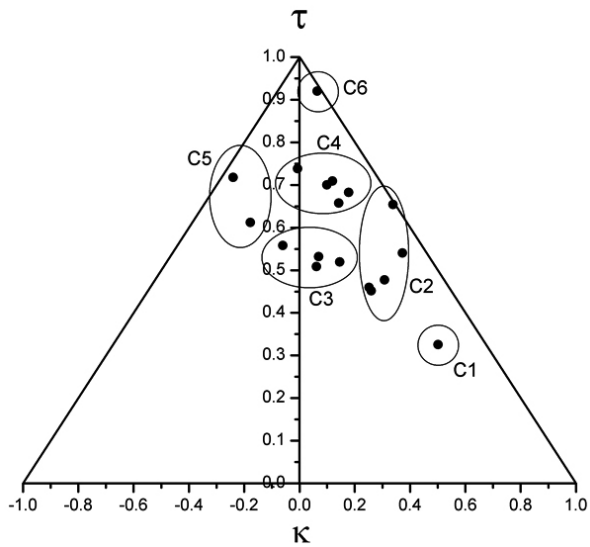

Figure 6. BMMS diagram for the octane isomers (see Table 5), where the clusters obtained using 'centroid' method and 'Euclidean' metric (see Figure 7 and Figure 8) are showed.

From the BMMS diagram of Figure 5, it is easy to show the regularity in the behaviour of τ and κ shape descriptors for homologous series of molecules. This is so clear for normal alkynes and normal acenes. The behaviour for normal alkanes is regular too if the even and odd cases are analysed separately. That effect is consequence of the successive addition of an eclipsed methyl group which changes the shape in alternate way. Normal acenes belong to the oblate region up to four benzene rings but then structures fall into the prolate region for molecules with five or more benzene rings. In general, the thickness decreases as the location of the molecule within normal alkane, normal alkyne and normal acenes homologous series increases. Normal alkynes are examples of 'perfect' prolate-box molecules. On the other hand, cyclobutane, hexamethylbenzene and coronene are examples of 'perfect' oblate-box molecules. The location of methane molecule on the top of the diagram (i.e. near to the cube-box point) is clear due its quasi-spherical molecular geometry.

Octane isomers family represents a simple but challenging set of molecules with branching diversity. From the BMMS diagram of Figure 6, it is easy to see that 78% of the eighteen octane isomers belong to the prolate zone. Normal octane is the most elongated prolate molecule (i.e. minor thickness and maximum asymmetry) whereas 2,2,3,3-tetramethyl-butane is the more compact molecule (i.e. maximum thickness and near zero asymmetry), and 2,3,4-trimethyl-pentane is the most symmetric molecule. The three oblate molecules are the ethyl substituted: 3-ethyl-hexane, 3-ethyl-2-methyl-hexane and 3-ethyl-3-methyl-hexane. Figure 7 shows a dendrogram with a cluster analysis using τ and κ BMMS descriptors. The cluster analysis was performed in Python language (from scipy.cluster.hierarchy import dendrogram, linkage) using 'centroid' method and 'Euclidean' metric in 'linkage' procedure. Six clusters were obtained in more or less certain correspondence of the intuitive three-dimensional visual perception of shape as Figure 8 shows. That clusters are also showed in Figure 6. Although BMMD methodology is a crude approximation, it allows the replacement of an intuitive perception of molecular shape by a quantitative description that preserves some elements of that natural approaching.

On the other hand, the correlation between BMMS descriptors (τ and κ) and properties (molecular and/or molar) is currently under study. Additionally, the introduction of new descriptors to improve the model in order to give a detailed description of the molecular inside the box is in progress.

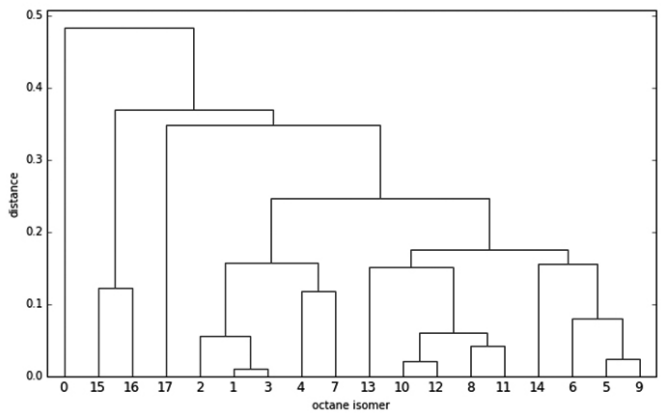

Figure 7. Similarity analysis of the octane isomers using the BMMS descriptors.

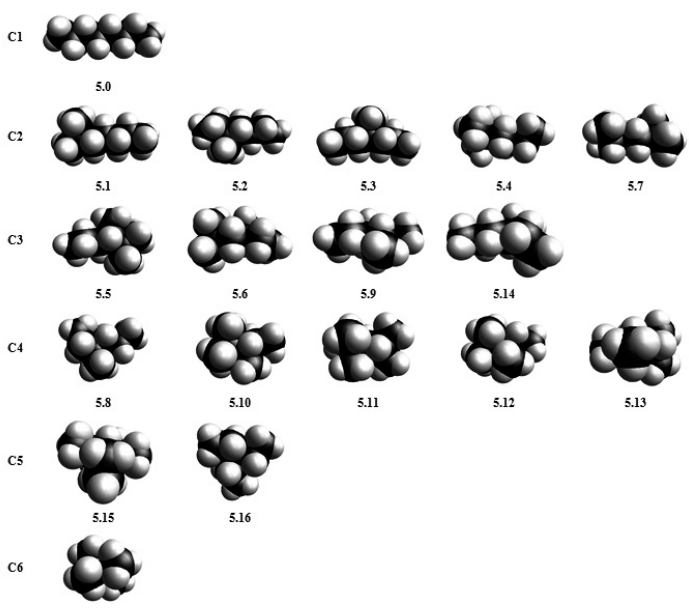

Figure 8. Clustering the 3D-vdW shapes of octane isomers by BMMS similarity analysis.

5. CONCLUSION

A simple model to molecular shape was introduced by two descriptors. Inspired into the old length-to-breadth ratio descriptor, the new parameters

result from the combination of length-width-height ratios of a rectangular box in which the molecule under study is enclosed. The similarity analysis of molecular shape for a set of molecules was facilitated by the introduction of a triangular diagram where the extreme cases (i.e. cubic-box, prolate-box and oblate-box) are clearly located. The present model is quite general to be applied to any molecule for which its molecular surface is known. Obviously, the next step for improving the present model is the introduction of extra parameters to give a more deeply description for the molecular surface morphology of the molecule inside the box, but the global molecular shape characterization given here by BMMS descriptors will remain invariant as the first-order approximation.

ACKNOWLEDGEMENTS

ARM thanks FONDECYT (Chile) under grant N°1080561.

REFERENCES

[1] P.G. Mezey, Shape in Chemistry. An Introduction to Molecular Shape and Topology, VCH Publishers, New York, 1993.
[2] A.Y. Meyer, Chem. Soc. Rev. 15 (1986) 449-474.
[3] A.Y. Meyer, Struct. Chem. 1 (1990) 265-279.
[4] Y. Marcus, J. Phys. Org. Chem. 16 (2003) 398-408.
[5] P.G. Mezey, Rev. Comput. Chem. 1 (1990) 265-294.
[6] A. Gavezotti, J. Am. Chem. Soc. 105 (1983) 5220-5225.
[7] M.L. Connolly, J. Am. Chem Soc. 107 (1985) 1118-1124
[8] K.D. Gibson, H.A. Scheraga, Mol. Phys. 62 (1987) 1247-1265.
[9] A.Y. Meyer, J. Chem. Soc. Perkin Trans. 2 (1985) 1161-1169.
[10] A. Bondi, Physical Properties of Molecular Crystals, Liquids, and Glasses, Wiley, New York, 1968.
[11] F.M. Richards, Ann. Rev. Biophys. Bioeng. 6 (1977) 151-176.
[12] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.
[13] D.G. Lister, J.N. Macdonald, N.L. Owen, Internal Rotation and Inversion. Academia Press, London, 1978.
[14] G.A. Arteca, Rev. Comput. Chem. 9 (1991) 191-253.
[15] J.E. Amoore, Ann. N. Y. Acad. Sci. 116 (1964) 457-476.
[16] P.A. Bath, A.R. Poirrette, P. Willett, F.H. Allen, J. Chem. Inf. Comput. Sci. 35 (1995) 714-716.
[17] D.D. Robinson, T.W. Barlow, W.G. Richards, J. Chem. Inf. Comput. Sci. 37 (1997) 939-942.
[18] N. Bodor, Z. Gabanyi, C.K. Wong, J. Am. Chem. Soc. 111 (1989) 3783-3786.
[19] A.Y. Meyer, J. Comput. Chem. 9 (1988)18-24.
[20] R. Kaliszan, H. Lamparczyk, A. Radecki, Biochem. Pharmacol. 28 (1979) 123-125.
[21] S.A. Wise, W.J. Bonnett, F.R. Guenther, W.E. May, J. Chromatogr. Sci. 19 (1981) 457-465.
[22] E.R. Collantes, W. Tong, W.J. Welsh, Anal. Chem. 68 (1996) 2038-2043.
[23] Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004.
[24] A. Pedretti, L. Villa, G. Vistoli, J. Mol. Graphics Modell. 21 (2002) 47-49. The executable and the source code of VEGA ZZ software can be free downloaded from http://www.vegazz.net