

ENTROPIC FACTOR AS A POTENTIAL MOLECULAR DESCRIPTOR

FACTOR ENTRÓPICO, UN DESCRIPTOR MOLECULAR POTENCIAL

YOANA PÉREZ-BADELL^{a,b} AND JOSÉ MANUEL NIETO-VILLAR^{a†}

a) Departamento de Química-Física, Cátedra Química "M.V. Lomonosov", Facultad de Química, Universidad de La Habana, La Habana 10400 Cuba. nieto@fq.uh.cu[†]

b) Laboratorio de Química Computacional y Teórica, Facultad de Química, Universidad de La Habana, La Habana 10400 Cuba.

† corresponding author

Received 24/03/2015; Accepted 01/06/2015

PACS: Entropy thermodynamics, 05.70.-a; Mixing fluids, 47.51.+a; Molecular structure, 33.15.-e.

The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment [1].

In this context, it is well known that many compounds show their active properties in solution. Nowadays, the development of molecular descriptors that take into account the effect of the solvent and the molecular properties of the compounds is essential in order to better determine the behavior in physiological systems.

Thermodynamics offers a formal framework for the development of molecular descriptors. For example, a set of molecular descriptors, based on Shannon entropy have been established: Shannon entropy (SE) [2], Differential Shannon entropy (DSE) [3–5], Shannon entropy descriptors (SHED) [6], extension of the Shannon entropy (DSE) [7] and a Mutual Conditional and Joint Entropy [8].

The aim of this communication is to develop a thermodynamic formalism to obtain a molecular descriptor that reflects the effect of the solvent based on the entropy of mixing.

In an ideal solution, the entropy of mixing [9] is given as:

$$\Delta S_M = -R \{n_1 \ln x_1 + n_2 \ln x_2\}, \quad (1)$$

where x_1 , x_2 represent the molar fractions of solvent and solute respectively.

In terms of molality m [10], Equation (1) may be rewritten as:

$$\Delta S_M = -R \left\{ \frac{10^3}{MM_1} \ln x_1 + m \ln x_2 \right\}, \quad (2)$$

where molar fractions are given by:

$$x_1 = \frac{n_1}{n_T} = \frac{\frac{10^3}{MM_1}}{\frac{10^3}{MM_1} + m} = \frac{k_1}{k_1 + m},$$

$$k_1 \gg m \therefore x_1 \cong 1,$$

$$x_2 = \frac{n_2}{n_T} = \frac{m}{k_1 + m} \cong \frac{m}{k_1}. \quad (3)$$

Substituting Equation (3) into (2), we obtain:

$$\Delta S_M = -R \left\{ m \ln \frac{m}{k_1} \right\} = -R \{m \ln m - m \ln k_1\}. \quad (4)$$

Note that expression (4) depends only on the molality of the solution.

An expression that takes into account the characteristics of the solute can be reached by the introduction of volume fractions ϕ_i , instead of molar fractions. Then, substitution into Equation (1) leads to the following:

$$\Delta S_M = -R \{n_1 \ln \phi_1 + n_2 \ln \phi_2\}, \quad (5)$$

where ϕ_1 y ϕ_2 are given as:

$$\phi_1 = \frac{\bar{V}_1}{V},$$

$$\phi_2 = \frac{\bar{V}_2}{V}, \quad (6)$$

and \bar{V}_1 , \bar{V}_2 denote the partial molar volumes [10] of solvent and solute, respectively. In order to take into account the molality, Equation (5) becomes:

$$\Delta S_M = -R \left\{ \frac{10^3}{MM_1} \ln \phi_1 + m \ln \phi_2 \right\} \quad (7)$$

Notice that equation (7) is given in terms of the volume fractions of solvent and solute respectively. In addition, for a dilute solution, Equation (7) as a function of the apparent molal volume φ_V [10] is given by:

$$\Delta S_M \approx -R \left\{ \frac{10^3}{MM_1} \ln \frac{V_1^0}{V} + m \ln \frac{\varphi_V}{V} \right\} \quad (8)$$

which can be rewritten as:

$$\Delta S_M = A + \Psi m \quad (9)$$

where A represents a constant, and a set of *entropic factors* (Ψ) characterizes the nature of the solute's interactions in solution. It represents a molecular descriptor.

Table 1. Van der Waals volume and entropic factor values as molecular descriptors used in the regression analysis.

N-acetylamino acid amides and oligopeptides	Molecular Formula	Van der Waals volume V_{vdW} in gas phase (\AA^3)	Entropic factor in aqueous solution Ψ
1. Glycine	$C_2H_5NO_2$	64.34	25.89303
2. Alanine	$C_3H_7NO_2$	81.33	23.14610
3. Serine	$C_3H_7NO_3$	88.14	23.21448
4. Acetylglycinamide	$C_4H_8N_2O_2$	103.98	19.97569
5. Threonine	$C_4H_9NO_3$	105.00	21.22885
6. Glycylglycine	$C_4H_8N_2O_3$	110.73	21.26559
7. Acetylanilinamide	$C_5H_{10}N_2O_2$	121.14	18.52713
8. Alanyl glycine	$C_5H_{10}N_2O_3$	127.81	19.59172
9. Histidine	$C_6H_9N_3O_2$	134.86	19.10213
10. Acetylvalinamide	$C_7H_{14}N_2O_2$	154.26	16.46126
11. Alanyl glycylglycine	$C_7H_{13}N_3O_4$	174.09	16.87439
12. Tetraglycine	$C_8H_{14}N_4O_5$	203.19	14.61049
13. Pentaglycine	$C_{10}H_{17}N_5O_6$	249.64	14.19300

Our approach has been tested on a set of thirteen different *N*-acetylamino acid amides and oligopeptides in aqueous solution. The experimental values of apparent molal volumes (φ_V) for different molalities (m) were taken from the literature [11, 12].

The van der Waals volume, that reflects the geometrical structure, is considered as a surface descriptor (V_{vdW}) [13]. Moreover, there is a strong dependence between the surface of the solute molecule and the solvation free energy [14].

V_{vdW} was determined by using fully optimized geometries based on Møller–Plesset second-order perturbation theory (MP2) [15] at the 6-31G (d) basis set level, using the Gaussian 09 program package [16].

Van der Waals molecular volume calculations were based on a grid method derived by Bodor *et al.* [17], using the atomic radii of Gavezzotti [18], included in HyperChem version 7.0 package [19]. Calculated properties of the molecules are shown in Table 1.

It shows how the *entropic factor* (Ψ) is sensitive with the *N*-acetylamino acid amides and oligopeptides in aqueous solution. Furthermore, Figure 1 presents the log-log plots of dependence between the entropic factor (Ψ) and the van der Waals volume (V_{vdW}) for the studied molecules. A power-law type dependence [20, 21] among molecular descriptors was found.

The correlation between both descriptors is expressed by the following relationship:

$$\Psi = \frac{k}{(V_{vdW})^\alpha} \quad (10)$$

An excellent correspondence between both descriptors is achieved. The linear fit $\ln \Psi = \ln k - \alpha \ln V_{vdW}$ was obtained

for $\ln k = 5.20$ and $\alpha = 0.5$.

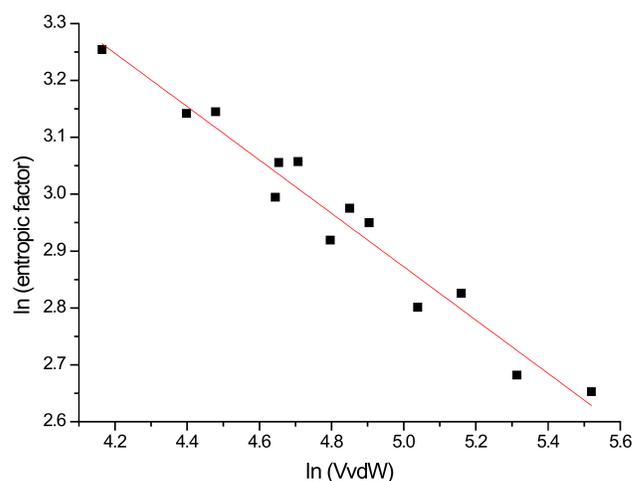


Figure 1. Power-law plot of the logarithm of entropic factor Ψ vs. logarithm van der Waals volume V_{vdW} for thirteen different *N*-acetylamino acid amides and oligopeptides in aqueous solution.

In conclusion, the entropic factor (Ψ), as a novel set of molecular descriptors based on the thermodynamics formalism of the entropy of mixing, was introduced. The approach has been tested on thirteen different *N*-acetylamino acid amides and oligopeptides in aqueous solution. High power-law type correlation between the entropic factor (Ψ) and the van der Waals volume (V_{vdW}) was observed, supporting the validation of the entropic factor as a molecular descriptor.

We thank Prof. Dr. A. Alzola *in memoriam*. Drs. C. Peniche, C. Perez and L.A. Montero for comments and suggestions. We acknowledge the Spanish Agency for International Cooperation for Development (AECID) for providing the resources that were used in the calculations. Finally, we thank the anonymous reviewers for their helpful comments and interesting suggestions.

[1] R. Todeschini and V. Consonni, Handbook of Molecular Descriptors, (Wiley-VCH Verlag GmbH: Weinheim, Germany, 2000).

[2] C. E. Shannon, Bell Syst. Technol. J. **27**, 379 (1948).

- [3] J. W. Godden, F. L. Stahura, J. Bajorath, *J. Chem. Inf. Comput. Sci.* **40**, 796 (2000).
- [4] J. W. Godden, J. Bajorath, *J. Chem. Inf. Comput. Sci.* **41**, 1060 (2001).
- [5] J. W. Godden, J. Bajorath, *J. Chem. Inf. Comput. Sci.* **42**, 87 (2002).
- [6] E. Gregori-Puigjané, J. Mestres, *J. Chem. Inf. Model.* **46**, 1615 (2006).
- [7] J. Batista, J. Bajorath, *J. Chem. Inf. Model.* **47**, 59 (2007).
- [8] J. Barigye Stephen, *et al.*, *J. Comput. Chem.* **34**, 259 (2013).
- [9] J. W. Gibbs, *Scientific Papers, Thermodynamics*, Vol. I, (Longmans Green: New York, 1906).
- [10] R. Quintana González, L. Giraldo Gutiérrez, J. C. Moreno Piraján and J. M. Nieto-Villar, *Termodinámica* (ed. Uniandes, 2005).
- [11] A. W. Hakin, G.R. Hedwing, *Phys. Chem. Chem. Phys.* **2**, 1795 (2000).
- [12] A. W. Hakin, H. Høiland, G. R. Hedwing, *Phys. Chem. Chem. Phys.* **2**, 4850 (2000).
- [13] Y. H. Zhao, M. H. Abraham, A. M. Zissimos, *J. Org. Chem.* **68**, 7368 (2003).
- [14] R. S. Pearlman, *Molecular Surface Area and Volume. In Partition Coefficient Determination and Estimation*, (Pergamon Press: New York, 1986).
- [15] C. Møller, M. S. Plesset, *Phys. Rev. A* **46**, 618, (1934).
- [16] M. J. Frisch, *et al.*, Gaussian 09, Revision A.01, Gaussian, Inc., Wallingford CT, (2009).
- [17] N. Bodor, Z. Gabanyi, C. K. Wong, *J. Am. Chem. Soc.* **111**, 3783, (1989).
- [18] A. Gavezzotti, *J. Am. Chem. Soc.* **105**, 5220, (1983).
- [19] HyperchemTM Release 7.0 for Windows, Molecular Modeling System, (Hypercube, Inc. 2002).
- [20] G. Nicolis, C. Nicolis, *Foundations of complex systems, Nonlinear Dynamics, Statistical Physics, Information and Prediction* (World Scientific Publishing Co. Pte. Ltd., 2007).
- [21] N. Boccara, *Modeling Complex Systems* (Springer-Verlag New York, Inc., 2004).