Modeling the Solubility of Aliphatic Alcohols *via* Molecular Descriptors

[1]G. P. ROMANELLI, [2]C. M. MARTINO AND [2]E. A. CASTRO*
[1]*Programa LADECOR, Departamento de Quimica, Facultad de Ciencias Exactas, UNLP, Calle 47 y 115, La Plata 1900, Argentina*
[2]*CEQUINOR, Departamento de Quimica, Facultad de Ciencias Exactas, UNLP, C.C. 962, La Plata 1900, Argentina*

(Received 8th June, 2000, revised 21st September, 2001)

Summary: We compare several QSPR models for predicting the aqueous solubility of aliphatic alcohols on the basis of physical (structural and electronic) descriptors. Results are compared with previously published data obtained from topological descriptors and with available experimental results. It is noted that there is a remarkable improvement on going from linear models in one variable to the quadratic equations in several variables. Some further possible extensions are discussed

Introduction

It is well known that alcohols are toxic substances and thus they constitute dangerous environmental pollutants especially in the case if an unfortunate accident takes place and large quantities of them enter accidentally within the ecological system. The first step in the polluting mechanism of alcohols is their solubility in water. The solubility of alcohols in water is a useful parameter in many applications including biochemical research. This is also the first step in the understanding alcohol transport in the living organisms.

Aqueous solubility is a physical property that has been extensively studied. As a property including water as solvent, it is significant in a diverse set and large number of situations, including chemical, biochemical, pharmaceutical, environmental, and industrial applications. Today there is no questions that the capability to predict the aqueous solubility of compounds is very useful.

Several different methods have been developed for the estimation of aqueous solubility with varying success and applicability [1-16]. The development of a quantitative structure-property relationship (QSPR) can be helpful to comprehend the aqueous solubility and can provide a simple enough method of estimating this property value directly from the chemical structure without being necessary to resort to any experimental measurement. The key aspect in a QSPR study is to develop a mathematical model which relates the structures of a set of compounds to a physical property such as aqueous solubility.

The fundamental assumption in a QSPR like study is that there are some sort of relationship between the chosen physical property of interest and the molecular structure. Several physical properties have been studied in this way and a QSPR may also lead to an understanding of the intrinsic structural features related to the physical property [17-19].

In a very recent article [20], Nikolic and Trinajstic have presented a comparison of several structure-property models for predicting the solubility of aliphatic alcohols in the water. They applied topological indices and the best model obtained was a two parameter model with ordered orthogonalized descriptors which includes explicitly the OH group. However, it is rather difficult to provide physical insight into the molecular behaviour for a purely topological representation of structure in many cases. Of late, more attention has been paid to the development and application of physical descriptors, such as electronic descriptors [21], geometrical descriptors [22,23] and semiempirical quantum chemical descriptors [10,11,24]. These descriptors can be helpful to understand the molecular structure, Therefore, we have deemed interesting and potentially useful to resort to some physical descriptors to model the aqueous solubility of aliphatic alcohols.

*To whom all correspondence should be addressed.

The aim of this study is to compare several QSPR models for predicting the aqueous solubility of aliphatic alcohols on the basis of physical descriptors.

QSAR/QSPR descriptors

It is well known that QSAR/QSPR are attempts to correlate molecular structure and properties derived from molecular structure with a particular class of chemical feature or biochemical activity. The kind of the feature or and activity is a function of the interest of the researchers and this spirit QSAR/QSPR is widely used in several investigation, development and applications fields, such as pharmaceutical, environmental, agricultural, physical, biological or chemical processes taking place in the target activity.

Although the set of QSPR descriptor is large enough and there are many possibilities to choose among of characteristic parameters which encompass several significant physical-chemistry properties. Thus, we have selected the log of the octanol-water partitions coefficient (Log P) which makes up a measure of hydrophobicily; polarizability (P); solvent-accessible surface-bounded molecular volume (SAG); mass (M); molar refractivity (RM) and volume (V) in orden to model the experimental solubilities in water of several alcohols in terms of this logarithms (Ln Sol).

Computations of solvent-accessible surface areas are carried out an approximate method due to Still and *et al.*, [25,26] . The procedure is fast, and usually accurate to within 10 percent for a given set of atomic radii very compact structures may present even greater errors than this one.

The polarizability is estimated from an additivity scheme presented by Miller [27] where different increments are associated with different atom types. For a wide and representative set of organic molecules, the estimates are accurate to around 1 to 3%. However, the method does not distinguish isomers of any type.

Calculation of Log P is carried out resorting to a set of atomic parameters derived by Ghose *et al.*, [28] and later on extended by Viswanadhan *et al.*, [29].

The refractivity is estimated *via* the same method as Log P. Ghose and Crippen presented atomic contributions to the refractivity in exactly the same manner as to the hydrophobicity [29,30].

The calculation of volume is very similar to the Surface Area computation and it employs a grid method described by Bodor *et al.*, [31] together with the atomic radii of Gavezotii [32].

Results and Discussion

We have divided the whole set of 54 aliphatic alcohols in two subsets; a) a training set of 44 molecules and b) a probe test of 9 molecules. This second set comprises the following molecules: 1) 2-pentanol, 2) 2-hexanol, 3) 3,3-dimethyl-2-butanol, 4) 2-methyl-2-hentanol, 5) 2,4-dimethyl-2-pentanol, 6) 3-methyl-2-hexanol, 7) 3-nonanol, 8) 2,3-din»ethyl-2-pentanol, 9) 1-hexadecanol while the remaining ones pertain to the first set.

The first set was used to find out the best regression equations. After a complete perusal, we obtained the following best one, two and three variables correlation's:

1- In (Sol) - 11.1068 -4.3842 10^{-2} SAG n = 44, EV- 0.2861, r^2 = 0.9867

2- In (Sol) = 9.1611 - 2.5256 10^{-2} V n = 44, EV = 0.2442, r^3 = 0.9889

3- In (Sol) = 3.2690 + 6.5729 10^{-3} V - 5.4530 10^{-5} V^2 + 2.9448 10^{-8} V^3 n - 44, EV = 0.2434, r^2 = 0.9894

4- In (Sol) = 12.4811 - 5.1269 10^{-2} SAG + 9.4292 10^{-6} SAG^2 n = 44. EV = 0.2908, r^2 = 0.9870

5- In (Sol) = 8.3582 - 2.1116 $10^{'2}$ V - 5.7988 10^{-1} Log P n =44, EV = 0.2364, r^2 = 0.9895

6- In (Sol) = 19.0856^{-1}.4481 10^{-1} SAG + 1.2154 P+ 1.6283 10^{-4} SAG^2-4.7649 10^{-2} P^2 n = 44, EV - 0.2298, r^2 = 0.9903

7- In (Sol) = -9.4768 + 2.2614 10^{-1} SAG - 6.6234 10^{-1} RM -9.1912 10-^{4} SAG^2

+ 2.4490 10-^{2} RM^2 +1.0468 10-^{6} SAG^3 - 2.9659 10^{-4} RM^3 n = 44. EV= 0.2356, r^2 = 0.9906

8- In (Sol) = -5.4067 - 3.7542 log P - 3.6601 RM +1.2125 M n-44,EV- 0.0915, r^2= 0.9960

Table 1: SAG, V, Log P, P, RM, Mass and In $(Sol)_{exp.}$ of 44 aliphatic alcohols

Alcohol	SAG	V	Log P	P	RM	Mass	In (Sol)exp
1-Butanol	251.94	348.23	0.94	8.75	22.13	74.12	0.09531
2-Butanol	247.55	344.91	0.96	8.75	21.95	74.12	0.06579
1 -Pentanol	281.60	401.41	1.34	10.59	26.74	88.15	-1.34707
3-Pentanol	273.15	392.64	1.43	10.59	26.48	88.15	-0.48613
2-Methyl-1 -butanol	268.75	389.56	1.34	10.59	26.61	88.15	-1.05840
3 -Methyl-1 -butanol	273.54	389.93	1.27	10.59	26.68	88.15	-1.16796
2-Methyl-2-butanol	266.07	383.33	1.04	10.59	26.59	88.15	0.33861
3-Methyl 2-butanol	269.30	385.05	1.36	10.59	26.42	88.15	-0.40497
1-Hexanol	312.98	455.53	1.73	12.42	31.34	102.18	-2.71810
3-Hexanol	306.26	446.43	1.82	12.42	31.08	102.18	-1.83258
3,3-Dimethyl-l-butanol	290.96	427.07	1.71	12.42	31.16	102,18	-2.59027
2,3-Dimethyl-2-butanol	292.28	430.59	1.68	12.42	31.15	102.18	-0.85097
2-Ethyl-l-butanol	288.65	430.29	1.74	12.42	31.21	102.18	-2.78709
4-Methyl-l-pentanol	303.32	443.40	1.67	12.42	31.28	102.18	-2.28278
3 -Methyl-2-pentanol	293.04	432.16	1.76	12.42	31.02	102.18	-1.63990
4-Methyl-2-pentanol	299.38	438.27	1.69	12.42	31.10	102.18	-1.81401
2-Methy 1-3-pentanol	291.80	434.04	1.83	12.42	30.95	102.18	-1.60944
3 -Methyl-3 -pentanol	286.68	427.75	1.51	12.42	31.11	102.18	-0.83011
1-Heptanol	342.35	508.64	2.13	14.26	35.94	116.20	-4.07454
3-Heptanol	335.69	499.42	2.22	14.26	35.68	116.20	-3.19418
4-Heptanol	336.59	500.42	2.22	14.26	35.68	116.20	-3.19662
2,4-Dimethyl-3-pentanol	311.71	473.87	2.23	14.26	35.42	116.20	-2.80181
2,2-Dimethyl-3-pentanol	307.49	470.06	2.34	14.26	35.35	116.20	-2.64367
2,3-Dimethyl-3-pentanol	305.76	467.78	1.91	14.26	35.59	116.20	-1.93794
2-Methyl-2-hexanol	328.11	491.34	1.83	14.26	35.79	116.20	-2.47337
1-Octanol	373.75	563.02	2.53	16.09	40.54	130.23	-5.40146
2-Octanol	370.46	559.98	2.54	16.09	40.36	130.23	-4.75600
2-Ethyl-l-hexanol	350.77	535.64	2.53	16.09	40.41	130.23	-4.99673
1-Nonanol	403.26	615.97	2.92	17.93	45.14	144.26	-6.90776
2-Nonanol	401.55	613.37	2.94	17.93	44.96	144.26	-6.31997
4-Nonanol	395.82	608.30	3.01	17.93	44.88	144.26	-5.95224
5-Nonanol	396.25	607.83	3.01	17.93	44.88	144.26	-5.74460
3,5,5-Trimethyl-l-hexanol	360.85	564.46	2.83	17.93	44.91	144.26	-5.76992
2,6-Dimethyl-4-heptanol	374.83	584.21	2.88	17.93	44.78	144.26	-5.77635
3,5 -Dimethyl-4-heptanol	358.77	566.85	3.03	17.93	44.62	144.26	-5.29832
7-Methyl-l-octanol	395.75	604.55	2.86	17.93	45.09	144.26	-5.74460
1-Decanol	434.80	670.25	3.32	19.76	49.74	158.28	-8.22080
1-Dodecanol	493.36	776.25	4.11	23.43	58.94	186.34	-10.6800
3-Ethyl-3-pentanol	309.25	469.46	1.97	14.26	36.64	116.20	-1.91732
2,2,3-Trimethyl-3-pentanol	317.34	499.03	2.41	16.09	39.98	130.23	-2.93182
2,2-Diethyl-l-pentanol	350.39	561.72	3.04	17.93	44.81	144.26	-5.57275
2,2-Dimethyl-l-propanol	266.68	384.11	1.45	10.59	26.40	88.1	-0.64626
l-Tetradecanol	556.88	888.52	4.81	27.10	68.14	214.39	-12.77171
1-Pentadecanol	587.00	938.54	5.30	28.94	72.75	228.42	-14.61402

9-In (Sol)=8.3741-2.9255 10^{-2} SAG-1.6388 log P + 3-9980 10^{-2} RM n = 44, EV " 0.2191, r^2 = 0.9905

The degree of agreement between theoretical and experimental data can be seen in Figure 1.

Then, the second set was employed for predictive purposes, and results are displayed in Table 2.

The analysis of precedent statistical data show that present structure-solubility models for aliphatic alcohols are really satisfactory. Besides, predictive results obtained for set 2 are good enough, with an acceptable average deviations for the different regression equations.

Experimental

We have made a complete regression analysis based on several variables (one, two, three, four and five) equations to first, second and third order, respectively. In a previous study on this issue for a more restricted set of saturated alcohols, we had shown the need to resort to several variables and

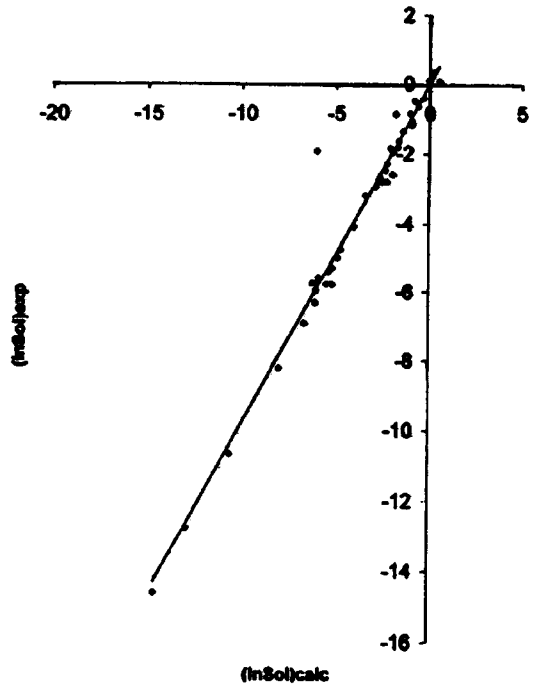

Fig. 1: A Plot of (ln Sol)exp vs. (ln Sol)calc computed using Eq. 8.

higher-order regression equations to obtain satisfactory enough results for this set of variables when studying toxicity of alcohols [33].

Regression analysis was developed *via* the Mathematica® for Windows, peruel Software [34] version 2.2 and molecular descriptors were calculated with Chem, Plus 1.6 (C P 10-6903-N-2708), extension for Hyperchem, molecular modeling for windows,

The experimental data for the aqueous solubility of aliphatic alcohols were taken from Amidon *et al.*, [35] and the Handbook of Chemistry and Physics [36].

In table 1, we give the experimental solubilities in water in terms of their logarithms [(In Sol)$_{exp}$] together with the chosen molecular descriptors for 54 aliphatic alcohols.

Conclusions

We have proposed several models to predict the aqueous solubility of aliphatic alcohols. The basic descriptors are related to the molecular structure and they complement other previous studies based on topological indices [20,37,38].

In particular, we have noted that there is a remarkable improvement on going from the linear models in one variable to the quadratic model in several variables. The statistical parameters corresponding to each regression equation show they are reliable enough to be employed as good predictors of the aqueous solubilities of aliphatic alcohols.

Table 2: The logarithm of experimental aqueous solubility (ln Sol)Expand the theoretical aqueous solubility (In Sol)Cacl.

Compound	(ln Sol) $_{Exp}$	(In Sol)calc					
		Eq.5	Eq.G	Eq.7	Eq.9	Eq.11	Eq.13
1	-0.63488	-1.22376	-0.91402	-0.91724	-0.85403	-1.23375	-0.80619
2	-1.9951	-2.48334	-2.27633	-2.20259	-2.21918	-2.41017	-2.09541
3	-1.41059	-1.64202	-1.62371	-1.57459	-1.65035	-1.50849	-1.98185
4	-1.1178	-1.91866	-1.89295	-1.83112	-1.71308	-1.81821	-1.04048
5	-2.14558	-2.71965	-2.93122	-2.85226	-2.77252	-2.74422	-1.93357
6	-2.26336	-2.82618	-2.99461	-2.91607	-2.90670	-2.84558	-2.34936
7	-6.1193	-6.19369	-6,18924	-6.26816	-6.22133	-6.22589	-6.05688
8	-2.002	-2.37330	-2.74584	-2.66656	-2.66391	-2.40142	-1.94109
9	-15.5874	-15.9696	-15.8926	-15.1247	-15.8940	-16.0607	-15.9438
r^2		0.9896	0.9891	0.9891	0.9895	0.9903	0.996

Conclusions derived form this study are that molecular descriptors are very good to be employed for this sort of study and they complement topological ones, with the extra advantage of giving the possibility to assign a structural and electronic interpretation to the statistical regression equations.

References

1. N. Bodor and M. Huang, *J. Pharm. Sci.*, **81**, 954 (1992).
2. G.Klopman, S.Wang and D. M. Balthasar, *J. Chem. Inf. Comput. Sci.* **32**, 474 (1992).
3. Huuskonen, M. Salo and J. Taskinen, *J. Pharm.*

Sci. **86**, 450 (1997).
4. G. S. Patil, *J. Hazard. Mater.* **36**, 35 (1994).
5. J. Nouwen and B. Hansen, *Quant. Struct. Act. Relat.* **15**, 17 (1996).
6. P. Ruelle and U. W. Kesselring, *J. Pharm. Sci.* **86**, 179 (1997).
7. P. Ruelle and U, W, Kesselring, *J. Chem. Soc., Faraday Trans.* **93**, 2049 (1997).
8. S. H. lalkowsky, *Chemosphere*, **26**, 1239 (1993).
9. E. Silla, I. Tunon, F. Villar and J. L. Pascual-Ahuir, *J. Mot. Struct. THEOCHEM* **254**, 369 (1992).
10. N. Bodor, A. Harget and M. Huang, *J. Am. Chem. Soc.* **113**, 9480 (1991).
11. N. Bodor, M. Huang and A. Harget, *Int. J. Quantum Chem: Quantum Chem. Symp.* **26**, 853 (1992).
12. T. M. Nelson and P.C. Jurs, *J. Chem. Inf. Comput. Sci.*, **34**, 601 (1994).
13. T. Suzuki, *J. Comput. Aided Mol. Design*, **5**, 149 (1991).
14. W. A. Li, W. J. Doncette and A. W. Andren, *Chemosphere*, **29**, 657 (1994).
15. S. H. lalkowsky, *Chemosphere*, **26**, 1239 (1993).
16. J. M. Sutter and P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **36**, 100 (1996).
17. B. E. Mitchell and P. C. Jurs, *J.Chem. Inf. Comput. Sci.* **38**, 489 (1998).
18. J. Huueskonen, M. Salo and J. Taskinen, *J. Chem. Inf. Comput. Sci.* **38**, 450 (1998).
19. C. Chenzhong and L. Zhillang, *J. Chem. Inf. Comput. Sci.* **38**, 1 (1998).
20. S. NikoliC and N. TrinajstiC, *SAR-QSAR Environ. Res.* **9**, 117 (1998).
21. K. Osmialowski, J. Halkicwicz, A. Radecki and R. Kaliszan, *J. Chromatogr.*, **346**, 53 (1985).
22. R. Rohrbangh and P. C. Jurs, *Anal. Chim. Acta* **199**, 99 (1987).
23. L. H. Hall and L. B. Kier, *J. Chem. Inf. Comput. Sci.* **35**, 1039 (1995).
24. N. Bodorand and M. Huang, *J. Pharm. Sci.*, **81**, 954 (1992).
25. W. Hasel, T. F. Hendrickson and W. C. Still, *Tet. Comput. Math.* **1**, 103 (1988).
26. W. C. Still, A. Tempczyk, R. C- Hawwiey and T. Hendrickson, *J. Am. Chem. Soc.* **112** 6127 (1990).
27. K. J. Miller, *J. Am. Chem. Soc.* **112**, 8533 (1990).
28. A. K. Ghose, P. Prichett and G. M. Crippen, *J. Comp. Chem.* **9**, 80 (1988).
29. V. N. Viswanadhan, A. K. Gnose, G. N. Revankarand R. K. Robins, *J. Chem. Inf. Comput. Sci.* **29**, 163 (1989).
30. A. K. Ghose and G. M. Crippen, *J. Chem. Inf. Comput. Sci.* **17**, 21 (1987).
31. N. Bodor, Z. Gabanyi and C. Wong, *J. Am. Chem. Soc.*, **Ill**, 3783 (1989).
32. A. Gavezotti, *J. Am. Chem. Soc.* **100** 5220 (1983).
33. G. P. Romaneili, L. F. R. Cafferata and E. A. Castro, *J. Mol. Struct. THEOCHEM* **504**, 261 (2000).
34. P. T. Tarn, "A Physicist's Guide to Mathematics®", Academic Press, New York, 1997.
35. G. L. Amidon, S. H. lalkowsky and S. Leung, *J. Pharm. Sci.* **63** 1858 (1974).
36. "Handbook of Chemistry and Physics", The 67th Edition, CRC Press, Boca Ralon.FL, 1987.
37. B. LuCiC, S. NikoliĆ, N. TrinajstiĆ, A. Juric and Z. MihaliĆ, Croat. *Chem. Acta*, **68** 417 (1995).
38. A. JuriC, M. Gagro, S. NikoliC and N. TrinajstiC, *J. Math. Chem.* **11**, 179 (1992).