ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2002) 277–280

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Using the theoretical linear energy solvation energy relationship to correlate and predict nasal pungency thresholds

George R. Famini [a,*], Denise Aguiar [b], Marvin A. Payne [b], Ryan Rodriquez [b], Leland Y. Wilson [b]

[a] *AMSSB-RAS-I, US Army Soldier and Biological Chemical Command, APG, MD 21010, USA*
[b] *Department of Chemistry, La Sierra University, Riverside, CA 92515, USA*

Received 31 March 2001; accepted 31 March 2001

Abstract

The theoretical linear solvation energy relationship (TLSER) has been used to correlate and characterize 44 nasal pungency threshold (NPT) values in man with parameters derived from semi-empirical molecular orbital theory. The resulting relationship provides good correlative ($R^2 > 0.92$) and predictive ($R^2_{cv} > 0.88$) capability. In addition, the TLSER parameters are used as a molecular probe to attempt to understand the fundamental properties influencing nasal pungency. © 2002 Elsevier Science Inc. All rights reserved.

Keywords: TLSER; NPT; VOC; QSPR; QSAR; LSER; LFER

1. Introduction

Volatile organic compounds (VOCs) are major contaminants, degrading air quality in both indoor and outdoor environments. Although, their lethal and mutagenic characteristics are of concern, simply sensory irritation is much more prevalent among these compounds, and one that causes a much greater problem to the population as a whole. In fact, the sensory irritation of VOCs is considered so great that threshold limits have been placed on this effect [1]. Berglund et al. identified several hundred VOCs as potential irritants in the non-industrial workplace [2]. Despite the importance and the potential health implications, sensory irritation have only been measured for 44 compounds to date [3].

Abraham and co-workers developed a multiple linear regression for human sensory irritation based on the linear solvation energy relationship approach (LSER) [4]. The LSER approach has been demonstrated by Abraham and by others to be an excellent relationship for correlating and characterizing biological properties in terms of fundamental physico-chemical descriptors [5–7]. The resulting LSER provided excellent correlative capability ($R^2 > 0.95$), and related nasal pungency thresholds (NPTs) to cavity, polarizability and hydrogen bonding effects. Abraham et al. have proposed to use this equation to predict NPT values for additional high impact VOCs [4].

It is to be noted that the LSER parameters of the Abraham approach are experimentally determined. This requires synthesis (of common materials) or purchase in order to determine the LSER parameters necessary to predict NPT (or other properties). Although, methods have been developed to predict the LSER parameters, this must be viewed as an estimation, rather than as a direct computation [8]. Therefore, using an estimated value (LSER parameters) to generate an estimated property (the NPT in this case) could result in the propagation of errors.

A much more intuitive approach would be to use descriptors directly in an LFER-based regression that could be computed from first principles. Patterned after the same methodology as the Abraham LSER, we developed the theoretical linear solvation energy relationship (TLSER) which combines the generalized linear solvation energy approach (shown in Eq. (1)) with semi-empirical molecular orbital theory [9–15].

$$\text{LOG property} = \text{bulk/cavity} + \text{dipolarity/polarizability} + \text{hydgrogen bonding} \quad (1)$$

Using the LSER philosophy as a philosophical underpinning, we systematically identified a set of parameters which replace, essentially one-for-one, the LSER parameters of Abraham. These parameters, as used in the TLSER, are computed solely from computational methods. Further, the TLSER uses descriptors from first principles, rather than as an estimation of an experimental property. The generalized

* Corresponding author. Tel.: +1-410-4362552; fax: +1-410-4365373.
E-mail address: george.famini@sbccom.apgea.army.mil (G.R. Famini).

1093-3263/02/$ – see front matter © 2002 Elsevier Science Inc. All rights reserved.
PII: S1093-3263(01)00124-3

TLSER is shown in Eq. (2).

$$\text{LOG property} = P_0 + aV_{mc} + b\pi_I + c\varepsilon_\alpha + d\varepsilon_\beta + eq_- + fq_+ \quad (2)$$

In this formalism, the V_{mc} is the molecular volume, and represents the size of the cavity necessary to solvate the molecule. The polarizability is depicted by polarizability index π_I, which is derived by dividing the polarization volume by the molecular volume. The resulting unitless term is a measure of the ability of the electrons to move throughout the molecule by an induced dipole. The hydrogen bonding acidity is represented by two quantities, a covalent acidity and an electrostatic acidity. The covalent acidity ε_α uses the E_{LUMO} in a linear transformation as shown in Eq. (3).

$$\varepsilon_\alpha = 0.30 - \frac{E_{LUMO} - E_{HOMO(H_2O)}}{100} \quad (3)$$

The electrostatic acidity is represented by the partial charge of the most positive hydrogen in the molecule. While this formalism works well for single hydrogen donor sites, or those molecules where multiple donor sites are close enough to appear as a single site, this method is incapable of handling multiple isolated hydrogen bonding acidity sites on the same molecule.

Similarly to the hydrogen bond acidity, the hydrogen bond basicity is represented by two terms, a covalent basicity terms ε_β, and an electrostatic basicity term q^-. The ε_β is a linear transformation of the E_{HOMO} of the solute, and is represented by Eq. (4).

$$\varepsilon_\beta = 0.30 - \frac{E_{LUMO} - E_{HOMO(H_2O)}}{100} \quad (4)$$

The ε_β is analogus to Eq. (3), but is calculated from the E_{LUMO} and $E_{HOMO(H_2O)}$. This linear transformation of the E_{LUMO} (for acidity) and E_{HOMO} (for basicity) was done for two reasons. First, it gives a scale with a common zero point for all compounds, the comparison to water. Second, the scale is monotonically increasing with increasing acidity (ε_α) or basicity (ε_β). In this way, the coefficients in the regressions are easier to interpret, and the signs of the coefficients immediately indicate the effect of acidity/basicity on the property of interest.

The q^- term is the absolute value of the most negative formal charge in the solute. In instances where the TLSER descriptors have been directly compared to the Abraham descriptors, both directly comparing the descriptors, and through correlation of the same properties, there has been a surprising correlation among the terms, indicating that where one set of descriptors correlate well to a given property, the other should as well [16].

The main purpose in developing the TLSER was to provide an a priori manner in which to provide chemically meaningful descriptors for the purposes of prediction. However, it was soon recognized that the TLSER descriptors also serve as a powerful probe to characterize solute/solvent interactions as well. This is consistent with the approach of Kamlet and Taft. In this way, the presence of a descriptor, and the magnitude and sign of the coefficient, can be used to provide a chemically meaningful characterization.

In recent papers, we have demonstrated the value of the TLSER in correlating and characterizing several in vitro biological activities [17–19]. In this paper, we use the TLSER to develop a multiple linear regression for the nasal pungency data of Cumetto–Muniz, and to compare that regression with the results of Abraham.

2. Calculational procedure

The experimental nasal pungency data of Cumetto-Muniz ($N = 43$) was used to develop the TLSER relationship. MOPAC v6.0 was used to optimize structure and generate the TLSER descriptors [20,21]. Systat and Codessa were used for all statistical analysis [22]. The correlation generated in this paper adheres to two conditions. One, the variance inflation factor (VIF) for each variable, which is a measure of intercorrelation of each independent descriptor against all other independent descriptor, is under 5.0 [23]. Second, each of the descriptors reported in the equation are significant at the 95% level as determined by the Student's *t*-statistic (*t*-stat).

3. Results and discussion

The compounds used in this study, along with the TLSER descriptors and the NPT values, are provided in Table 1. Using this data, Eq. (5) shows the resulting TLSER regression.

$$\text{LOG } 1/\text{NPT} = 2.258V_{mc} + 3.471q_- + 50.56\varepsilon_\alpha + 13.63q_+ - 16.40 \quad (5)$$

t-Stat	17.1	4.30	7.38	13.1	12.9
VIF	1.19	1.19	2.16	2.05	

$N = 42$, $R^2 = 0.923$, standard deviation (S.D.) $= 0.357$, $F = 110$, $R^2_{cv} = 0.885$.

Oct-1-yne and ethanoic acid were removed from the regression as outliners (residual of greater than 2 S.D.). The correlative ability of the TLSER can be immediately seen by the excellent R^2 of 0.923, and the low S.D. The cross validated R^2 (R^2_{cv}) shows the predictivity of the regression. The value of 0.885 demonstrates an excellent fit. Fig. 1 shows a plot of the predicted versus the observed values, and highlights the even nature of the data and the predicted results. The variance inflation factor (VIF) in the above equation is the degree of correlation of each descriptor against all other independent descriptors. A value under 5.0 is normally considered acceptable, indicating no cross correlation. VIF or tolerance is generally considered superior to a correlation matrix as the latter considers only

Table 1
TLSER descriptors for the data set of Cometto-Muniz

Name	V_{MC}	π_i	ε_β	ε_α	q_-	q_+	NPT
Oct-1-yne	1.360	0.107	0.139	0.160	0.179	0.157	4.49
Propanone	0.631	0.100	0.138	0.172	0.285	0.023	5.12
Pentan-2-one	0.941	0.107	0.139	0.171	0.284	0.023	3.47
Heptan-2-one	1.295	0.107	0.139	0.171	0.286	0.03	2.91
Nonan-2-one	1.658	0.106	0.139	0.171	0.288	0.031	2.53
Methyl acetate	0.660	0.107	0.131	0.169	0.357	0.026	5.05
Ethyl acetate	0.819	0.111	0.131	0.169	0.357	0.026	4.83
Propyl acetate	1.024	0.107	0.132	0.169	0.358	0.026	4.24
Butyl acetate	1.238	0.104	0.132	0.169	0.358	0.027	3.56
s-Butyl acetate	1.207	0.106	0.133	0.168	0.361	0.026	3.60
t-Butyl acetate	1.218	0.105	0.133	0.168	0.362	0.025	3.98
Pentyl acetate	1.342	0.110	0.132	0.169	0.357	0.026	3.22
Hexyl acetate	1.520	0.110	0.132	0.169	0.357	0.026	2.80
Octyl acetate	1.865	0.110	0.135	0.168	0.329	0.037	1.95
Decyl acetate	2.289	0.106	0.132	0.169	0.358	0.026	0.70
Dodecyl acetate	2.547	0.110	0.133	0.172	0.358	0.032	0.10
Methanol	0.328	0.095	0.132	0.142	0.333	0.182	4.53
Ethanol	0.514	0.098	0.133	0.143	0.324	0.18	3.91
Propan-1-o1	0.688	0.101	0.133	0.144	0.325	0.18	3.49
Propan-2-ol	0.681	0.101	0.133	0.145	0.320	0.178	4.26
Butan-1-ol	0.841	0.105	0.133	0.146	0.323	0.179	3.20
s-Butanol	0.872	0.101	0.134	0.145	0.319	0.179	3.76
t-Butanol	0.842	0.104	0.134	0.144	0.318	0.176	4.52
Pentan-1-ol	1.024	0.104	0.133	0.146	0.325	0.18	3.21
Heptan-1-o1	1.330	0.109	0.133	0.147	0.325	0.18	2.32
Hexan-1-ol	1.241	0.101	0.133	0.146	0.325	0.18	2.62
Heptane-4-ol	1.388	0.104	0.135	0.148	0.321	0.179	2.53
Octan-1-ol	1.520	0.108	0.133	0.147	0.325	0.18	1.99
Toluene	0.964	0.128	0.153	0.176	0.101	0.06	4.47
Ethylbenzene	1.143	0.124	0.153	0.176	0.088	0.06	4.00
Propylbenzene	1.323	0.122	0.153	0.176	0.089	0.06	3.17
Chlorobenzene	0.957	0.129	0.149	0.179	0.112	0.078	4.02
Pyridine	0.760	0.125	0.149	0.178	0.230	0.084	3.11
Butanal	0.782	0.105	0.138	0.170	0.286	0.033	4.77
Pentanal	0.957	0.106	0.138	0.170	0.286	0.032	4.57
Hexanal	1.127	0.106	0.138	0.170	0.286	0.032	3.70
Heptanal	1.280	0.109	0.138	0.170	0.286	0.032	3.13
Octanal	1.492	0.106	0.138	0.170	0.286	0.032	3.24
Formic acid	0.328	0.095	0.128	0.168	0.369	0.216	2.50
Ethanoic acid	0.488	0.103	0.132	0.169	0.305	0.2	1.62
Butanoic acid	0.841	0.105	0.131	0.169	0.366	0.216	1.79
Hexanoic acid	1.180	0.107	0.131	0.169	0.366	0.216	1.30
Octanoic acid	1.560	0.105	0.131	0.169	0.366	0.216	0.30

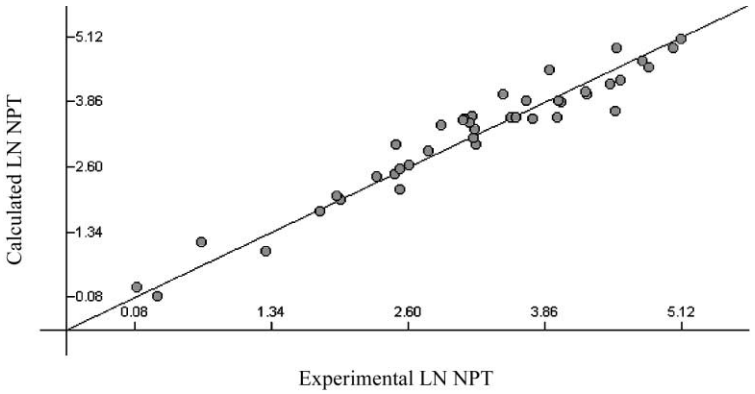

Fig. 1. Experimental vs. predicted plot for Eq. (4).

two-fold interactions [23]. The above VIFs show that there is no cross correlation between descriptors.

Using the coefficients as a probe, it is possible to highlight characteristics of compounds with high sensory irritation. The most significant coefficient in the regression is the volume term, with a *t*-statistic of 17.1 (shown in the *t*-stat line of Eq. (5)). This is indicative of cavity or dispersion effects, and has been found to be significant when transport is a key factor. The second highest *t*-statistics are those corresponding to the two acidity terms, q_+ and ε_α. This suggests that the receptor region for nasal pungency contains a basic site. Because the electrostatic term is much more significant, it is likely that the basic site is one or more small heteroatoms, such as N or O. However, the significance of the covalent term could indicate the addition of a larger heteroatom (such as S) or some conjugation around the active site.

The significance of the electrostatic basicity, like the volume term, is consistent with the transport phenomena found in previous TLSERs. Although this could also be indicating the presence of permanent dipolar effects (coupled with the q_+) term, the lack of a polarizability term, and the overwhelming significance of the acidity term makes this unlikely. A principal component analysis of the descriptors did not result in any of the principal components containing both the q_- and q_+ with approximately equal contributions to an eigenvector, further confirming that dipolar effects are probably not significant in nasal pungency.

Because, both the Abraham regression and the TLSER are based on the general linear solvation energy relationship, they can be directly compared. The Abraham regression is shown in Eq. (6) [4].

$$\text{LOG}\,(1/\text{NPT}) = 2.154\pi_2^{\text{H}} + 3.522\sum\alpha_2^{\text{H}} + 0.860\,\text{LOG}\,L^{16} - 8.519 \quad (6)$$

$N = 43$, $R^2 = 0.955$, S.D. $= 0.27$, $F = 201$.

As can be seen from comparing the statistics of Eqs. (5) and (6), the goodness of fit are about equivalent, with the Abraham LSER slightly better. Although, Abraham does not report *t*-statistics, the LSER parameters used are usually in the same range, making a general comparison of coefficients possible, at least in a qualitative sense. From Eq. (6), the acidity term, $\sum\alpha_2^{\text{H}}$ is far the most significant, indicating the likelihood of a basic site in the receptor region for nasal pungency. Again, this is entirely consistent with the findings of the TLSER. Further, the presence of the L^{16} and π_2^{H} descriptors are indicative of dispersion interactions. The coefficients in both equations are positive, and the key interactions are the same, with acidity being indicated as the key driver, followed by dispersion/cavity interactions, and finally basicity.

References

[1] Y. Alaire, Environ. Health Persect. 42 (1981) 9.
[2] B. Berglund, U. Berglund, R. Lindvall, in: Proceedings of the ASHRAE Conference on IAQ, 1986, 138 pp..
[3] J.E. Cometto-Muniz, W.S. Cain, Arch. Occup. Environ. Health 71 (1998) 105.
[4] M.H. Abraham, R. Kumarsingh, J.E. Cometto-Muniz, W.S. Cain, Arch. Toxicol. 72 (1998) 227.
[5] M.H. Abraham, Chem. Soc. Rev. 22 (1993) 73.
[6] M.J. Kamlet, R.W. Taft, G.R. Famini, R.M. Doherty, Acta Chem. Scand. 41 (1987) 589.
[7] R.W. Taft, M.H. Abraham, G.R. Famini, R.M. Doherty, J.L. Abboud, M.J. Kamlet, J. Pharm. Sci. 74 (1985) 807–814.
[8] J.A. Platts, D. Butina, M.H. Abraham, A. Hersey, J. Chem. Inf. Comp. Sci. (1999) 39.
[9] G.R. Famini, Using theoretical descriptors in quantitative structure activity relationships, Part V, US Army Edgewood Research Development and Engineering Center: Aberdeen Proving Ground, Vol. CRDEC-TR-095, MD, 1989.
[10] G.R. Famini, C.E. Penski, L.Y. Wilson, J. Phys. Org. Chem. 11 (1992) 1992.
[11] G.R. Famini, L.Y. Wilson, J. Phys. Org. Chem. 6 (1993) 539.
[12] G.R. Famini, L.Y. Wilson, Using theoretical descriptors in LSER, in: G.R. Famini, L.Y. Wilson (Eds.), Vol. 1, Elsevier, 1995, 213 pp.
[13] G.R. Famini, L.Y. Wilson, Chemosphere 1997.
[14] G.R. Famini, L.Y. Wilson, J. Phys. Org. Chem. 12 (1999) 1.
[15] W.H. Donovan, G.R. Famini, J. Chem. Soc. Perkins 2 (1996) 83.
[16] J.S. Murray, P. Politzer, G.R. Famini, J. Mol. Struct. Theochem. 454 (1998) 299.
[17] G.R. Famini, V.P. Loumbev, E.K. Frykman, L.Y. Wilson, Quant Struct.-Act Relat. 17 (1998) 558.
[18] G.R. Famini, W.P. Ashman, A.P. Miskiewicz, L.Y. Wilson, QSAR 11 (1992) 162.
[19] L.Y. Wilson, G.R. Famini, J. Med. Chem. 34 (1991) 1668–1674.
[20] J.J.P. Stewart, MOPAC Manual Frank J. Seiler Research Laboratory, Vol. FJSRL-TR-88-007, 1988.
[21] M.J.K. Dewar, W. Thiel, J. Am. Chem. Soc. 99 (1977) 4899.
[22] Codessa Semichem, Inc., Sewanee, KS.
[23] D.A. Belesley, E. Kuh, R.E. Welsh, Regression Diagnostics, Wiley, New York, 1980.