JOURNAL OF PHYSICAL ORGANIC CHEMISTRY
J. Phys. Org. Chem. **12**, 645–653 (1999)

Using theoretical descriptors in linear free energy relationships: characterizing several polarity, acid and basicity scales

George R. Famini[1]* and Leland Y. Wilson[2]

[1]International Programs Office, US Army Soldier and Biological Chemical Command, Aberdeen Proving Ground, Maryland 21010, USA
[2]Department of Chemistry, La Sierra University, Riverside, California 92515, USA

Received 18 December 1998; revised 22 February 1999; accepted 2 March 1999

ABSTRACT: The widespread application of computational techniques to studies in biological, chemical and environmental sciences has led to a quest for important, characteristic molecular parameters that may be directly derived from these computational methods. Theoretical linear solvation energy relationships (TLSER) combine computational molecular orbital parameters with the linear solvation energy relationship (LSER) of Kamlet and Taft to characterize, understand and predict properties which depend upon solute–solvent interactions. The TLSER methodology was used to correlate and to attempt to understand and probe eight solvent scales commonly used in linear free energy relationships and physical organic chemistry. Correlations are presented which demonstrate the ability of the TLSER not only to predict values for these solvent scales, but which also serve as a probe to attempt to understand the underlying physical meaning of these solvent scales. This article is a US Government work and is in the public domain in the United States.

KEYWORDS: linear free energy relationships; theoretical descriptors; solvent scales

INTRODUCTION

Inherent to chemistry is the concept that there is a relationship between the macroscopic (bulk) properties of chemical compounds and their microscopic (structural) features.[1–3] Quantitative structure activity–property relationships (QSAR/QSPR) and linear free energy relationships (LFER) attempt to quantify, usually in the form of a linear or a linearizable regression, the relationship between structure and function.

Solute–solvent properties have generated a great deal of interest in the computational chemistry community in recent years.[4,5] In particular, the interaction between the solvent and the solute including the effect of solvent on solute molecular structure and, therefore, the bulk properties of the solute is being studied through a wide variety of techniques. There are three basic methods which one can use to study solute–solvent interactions. The most rigorous is the explicit model case, where all solvent molecules are treated individually and explicitly.[6] This is by far the most time consuming, and although there are cases of its use in quantum chemical studies, it has been used most frequently in molecular dynamics and Monte Carlo simulations. At the same time, this method generates the most information, not only structure deformities, but also explicit solute–solvent and solvent–solvent bonding.

The second method is the implicit model, where the solute is treated explicitly, but a continuum model or potential is used to portray the solvent.[7] Although this does not generate specific solute–solvent or solvent–solvent bonding, it does generate free energies of solvation, and is rapid enough to permit computation using *ab initio* or semi-empirical methods. This model has generated the most interest in the past several years.

The third method for studying solute–solvent interactions is the extension of LFER or QSAR to the concept of solvation, and represents the most empirical of the three methods.[3,8–16] A relatively large set of property data is required in order to generate a relationship. The advantage of this methodology is that relatively little information is needed about the correlated property/system, and often very complex 'solvent systems' such as receptors may be characterized. The major disadvantage is that only inferences about the solvent system results can be made, and only information about the particular information is generated. Further, the resulting correlation is valid for the interpolation only, and may not be valid for the extrapolations outside of the data set region. Still, for complex systems such as receptor sites, where it is unlikely that either of the first two methods can be readily used, the QSAR/QSPR/LFER methods combined

**Correspondence to:* G. R. Famini, International Programs Office, US Army Soldier and Biological Chemical Command, Aberdeen Proving Ground, Maryland 21010, USA.

CCC 0894–3230/99/080645–09 $17.50

with theoretical descriptors can give useful insights into important features. This paper will focus on the use of this third methodology, LFER, QSAR and relevant subsets of these two methods, to study solute–solvent interactions.

Quantitative structure–activity relationships, as originally developed by Hansch, were intended to aid medicinal and pharmaceutical chemists in identifying new drugs from biological activity data. An obvious extension was to examining not just drug activity but drug (chemical) toxicity.[17] Researchers have employed Hansch's QSAR approach to correlate different types of tonicities of compounds. Using the general QSAR methodology, a wide range of theoretical descriptors have been identified that correlate well with a variety of biological activities.[8,13,18–20]

A major step forward in applying LFER to solute–solvent interactions was the work of Kamlet and Taft, who developed an empirical molecular parameter approach to characterize a wide number of solute–solvent interactions in a subset approach of linear free energy relationships (LFER).[21–24] This approach, termed the linear solvation energy relationship (LSER), has been used successfully to characterize a wide range of physical, chemical and biological properties which are affected by, or which are, solute–solvent interactions. The generalized approach is shown in Eqn. (1), where any solute–solvent property may be divided into three separate effects:

$$\begin{aligned}\log(\text{property}) = {} & \text{bulk/cavity} \\ & + \text{dipolarity/polarizability} \\ & + \text{hydrogen bonding}\end{aligned} \tag{1}$$

Most solute–solvent properties follow this construct, and can be expressed as a linear combination of these terms. The bulk/cavity term describes the energy required to break the solvent matrix and insert the solute. The dipolarity/polarizability term is the interaction between the induced dipoles of the solute and solvent molecule. The hydrogen bonding describes the acidity and basicity of the solute and solvent and the interaction between the two terms. Kamlet and Taft,[22] Abraham and co-workers,[25–28] and Carr *et al*. have each developed a set of empirically based parameters for each of these interaction terms which have been used to successfully correlate a wide variety of physical, chemical and biological properties. Abraham has recently extended the LSER by using gas chromatography instead of solvatochromic shifts to determine the LSER parameters, and added a dispersion term, which is modeled by the gas–hexadecane partition coefficient.

Ford and Livingstone[29] pointed out that there are certain advantages, and indeed opportunities, for using computationally or theoretically derived parameters instead of empirically derived parameters. Further, parameters from molecular orbital-based methods tend to have the added advantage of not being class dependent, as is often the case with group theoretical, topological or topographical indices. MO-derived descriptors are also readily computed, and can be selected in such a way as to ensure (or nearly ensure) orthogonality of the parameter space.[18,19] There are currently a number of methodologies in the literature for examining physical, chemical and biological properties using computationally derived descriptors.[13–15,19,30–32]

Following this philosophy, we have endeavored to utilize the generalized LSER as a framework and use computationally derived descriptors instead of the empirical descriptors of the LSER. The theoretical linear solvation energy relationship (TLSER) was developed with the express aim of maintaining the function of the LSER, but using a one-for-one (as much as possible) replacement of empirical descriptors with computational descriptors. The resulting TLSER is

$$\begin{aligned}\log(\text{property}) = {} & aV_{\text{mc}} + b\pi_{\text{I}} + d\varepsilon_{\beta} + eq_{-} + c\varepsilon_{\alpha} \\ & + fq_{+} + g\end{aligned} \tag{2}$$

In this formalism, V_{mc} represents the molecular volume, and depicts the size of the 'hole' necessary in the solvent matrix. In addition, when used for solvents, V_{mc} is inversely proportional to the cohesion energy, usually depicted as the Hildebrand solubility parameter. The polarizability term, π_{I}, is described by dividing the polarization volume by the molecular volume. This results in a unitless term independent of volume that depicts the ease with which electrons may move throughout the molecule. The hydrogen bonding term basicity is represented by two terms, and covalent basicity and an electrostatic basicity. The covalent basicity (ε_{β}) is defined as a linear transformation of the E_{HOMO}, specifically 0.30–0.01 [$E_{\text{LUMO(water)}} - E_{\text{HOMO}}$]. This transformation provides a 'zero point' reference for the scale, and corrects the scale to be positive for increasing basicity. The electrostatic basicity (q_{-}) is taken as the absolute value of the most negative formal charge (non-hydrogen) in the molecule. The hydrogen bond acidity is likewise divided into a covalent and electrostatic components. The covalent acidity (ε_{α}) is identical with ε_{β}, but uses the HOMO of water and the LUMO of the substrate. The electrostatic acidity (q_{+}) is defined as the value of the most positive hydrogen atom in the molecule. This methodolgy has been used to successfully develop correlation equations for a wide variety of properties.[11,12,33–42]

An assumption of this method, and indeed most LFER/QSAR methods involving computationally derived descriptors, is that the descriptors are 'pure' and reflect a particular microscopic property without 'mixing' or contamination from other descriptors. While impossible to prove in a 'global' sense, most of the TLSER

Table 1. Solvent scales used in the study

Parameter	Name	
AN	Acceptor number	(Gutmann)
DN	Donor number (donicity)	(Gutmann)
E^{T}_{N}	Solvatochromic polarity (normalized)	(Reichardt)
Y	Polarity function	(Koppel Palm)
P	Polarizability	(Koppel Palm)
MR	Molar refraction	(Koppel palm)
B	Basicity	(Swain)
E	Acidity	(Swain)
V_{mc}	Molecular volume	(TLSER)
π_I	Polarizability index	(TLSER)
ε_B	Covalent HBAB	(TLSER)
q_-	Electrostatic HBAB	(TLSER)
ε_A	Covalent HBDA	(TLSER)
q_+	Electrostatic HBDA	(TLSER)

regressions to date have very small cross-correlations among descriptors. This indicates that for the local data sets for which the regressions were computed there is little or no 'mixing' or cross-contamination of descriptors.

Although derived differently, with a different focus, the GIPF methodology of Politzer and Murray[43–45] has been shown also to be directly related to the LSER and to the TLSER. Based on electrostatic potential surfaces, this method is the *ab initio* alternative to the semi-empirical-based TLSER and the empirical LSER. Where comparisons have been made between the GIPF and the TLSER, the conclusions have generally been equivalent and the regressions comparable.[46]

In a recent paper, we correlated the effect of 30 solvents on the C=O stretching frequencies of several substituted 2-pyrrolidinones using the TLSER and five additional experimentally derived solvent scale methodologies.[47] In performing these correlations, a question arose regarding the information contained within the empirical scales. This paper compares several of these scales, in particular the Reichardt E^{T}_{N}, the Gutmann *AN* and *DN* and the Koppel Palm *Y*, *E*, *B* and *A* with the TLSER. In a related series of papers, Katritzky *et al.*[48] used semi-empirical molecular orbital methods to develop correlations for 45 solvent scales, including the Gutmann, Reichardt and Koppel Palm scales. Using CODESSA, Katritzky *et al.*[49] identified regression equations using about 120 different descriptors among the 45 correlated equations. Although the methodology used in CODESSA is very valuable in providing the 'best' correlation for a given property, having 118 different descriptors for the 45 scales makes any sort of comparison based on the computed and correlated computational descriptors very difficult.

Table 2. TLSER parameters

No.	Compound	V_{mc}	π_I	ε_β	q_-	ε_α	q_+
1	Cyclohexane	1.0643	0.1055	0.1278	0.0101	0.1461	0.0051
2	Hexane	1.1988	0.0993	0.1248	0.0218	0.1450	0.0045
3	Heptane	1.3652	0.1010	0.1254	0.0217	0.1455	0.0045
4	Triethylamine	1.3301	0.1009	0.1507	0.4324	0.1520	0.0055
5	Tetrachloromethane	0.9058	0.1172	0.1128	0.0703	0.1911	0.0000
6	Butoxybutane	1.6382	0.1014	0.1361	0.3452	0.1478	0.0177
7	Toluene	1.0192	0.1207	0.1524	0.1007	0.1756	0.0598
8	Benzene	0.8457	0.1205	0.1513	0.0593	0.1744	0.0593
9	Ethoxyethane	0.9046	0.0995	0.1361	0.3423	0.1455	0.0071
10	Chlorobenzene	0.9951	0.1241	0.1490	0.1118	0.1794	0.0777
11	Tetrahydrofuran	0.7895	0.1020	0.1374	0.3277	0.1471	0.0209
12	Bis(2-methoxyethyl) ether	1.4066	0.1034	0.1355	0.3579	0.1502	0.1280
13	Trichloromethane-*d*	0.7540	0.1114	0.1160	0.1122	0.1849	0.0876
14	Dichloromethane	0.6046	0.1036	0.1203	0.1602	0.1773	0.0555
15	Pyridine	0.7936	0.1200	0.1483	0.2299	0.1780	0.0835
16	Nitrobenzene	1.0079	0.1316	0.1421	0.3288	0.1860	0.0851
17	1,2-Dichlorobenzene	0.7858	0.1046	0.1210	0.1967	0.1789	0.0489
18	Cyanobenzene	0.9969	0.1276	0.1471	0.0867	0.1833	0.0696
19	Propanone	0.6401	0.0980	0.1376	0.2847	0.1715	0.0234
20	1,4-Dioxane	0.8547	0.1050	0.1387	0.3230	0.1478	0.0327
21	2-Methyl-2-propanol	0.8943	0.0975	0.1338	0.3180	0.1442	0.1764
22	Dimethyl sulfoxide	0.7219	0.1045	0.1471	0.7204	0.1734	0.0500
23	Cyanomethane	0.4529	0.0937	0.1173	0.1146	0.1622	0.0209
24	Nitromethane	0.4740	0.1092	0.1298	0.3348	0.1818	0.0498
25	2-Propanol	0.7144	0.0962	0.1331	0.3200	0.1450	0.1781
26	Butanol	0.9082	0.0969	0.1322	0.3249	0.1449	0.1804
27	Ethanoic acid	0.5249	0.0956	0.1294	0.3651	0.1696	0.2161
28	Ethanol	0.5435	0.0924	0.1322	0.3234	0.1429	0.1799
29	Methanol	0.3647	0.0860	0.1310	0.3291	0.1402	0.1803
30	Deuterium oxide	0.1782	0.0630	0.1233	0.3256	0.1237	0.1628

Table 3. Solvent scales

No.	Compound	*AN*	*DN*	E^{T}_{N}	*Y*	*P*	*MR*	*B*	*E*	π^*	β	α
1	Cyclohexane	—	0.0	0.006	0.254	0.256	27.2	0	0.000	0.00	0.00	0.00
2	Hexane[a]	0.0	0.0	0.009	0.235	0.236	34.5	0	0.06	−0.08	0.00	0.00
3	Heptane	0.0	0.0	0.012	0.235	0.236	34.5	0	0.00	−0.08	0.00	0.00
4	Triethylamine	1.4	61.0	0.043	0.321	0.243	33.8	650	0.00	0.14	0.71	0.00
5	Tetrachloromethane	8.6	0.0	0.052	0.292	0.274	26.4	00	0.00	0.28	0.00	0.00
6	Butoxybutane	—	19.0	0.071	0.407	0.242	41.0	285	0.000	0.24	0.46	0.00
7	Toluene	—	0.1	0.099	0.315	0.293	31.1	58	1.30	0.54	0.11	0.00
8	Benzene	8.2	0.1	0.111	0.300	0.295	26.3	48	2.10	0.59	0.10	0.00
9	Ethoxyethane	3.9	19.2	0.117	0.526	0.217	22.4	280	0.00	0.27	0.47	0.00
10	Chlorobenzene	—	3.3	0.188	0.606	0.306	31.0	38	0.00	0.71	0.07	0.00
11	Tetrahydrofuran	8.0	20.0	0.207	0.681	0.247	20.0	287	0.00	0.58	0.55	0.00
12	Bis(2-methoxyethyl) ether	10.2	20.0	0.231	0.667	0.231	24.0	238	0.00	0.53	0.41	0.00
13	Trichloromethane-*d*[b]	23.1	4.0	0.256	0.559	0.252	21.3	14	3.28	0.58	0.00	0.44
14	Dichloromethane	20.4	1.0	0.269	0.729	0.256	16.4	23	2.70	0.82	0.00	0.185
15	Pyridine	14.2	33.1	0.302	0.790	0.299	24.2	472	0.00	0.87	0.64	0.00
16	Nitrobenzene	14.8	4.4	0.324	0.918	0.322	32.8	67	0.00	1.01	0.39	0.00
17	1,2-Dichloroethane	16.7	0.0	0.327	0.757	0.266	21.0	40	3.00	0.81	0.00	0.00
18	Cyanobenzene	15.5	11.9	0.333	0.890	0.308	31.4	155	0.00	0.90	0.41	0.00
19	Propanone	12.5	17.0	0.355	0.868	0.220	16.2	224	2.10	0.71	0.48	0.80
20	1,4-Dioxane	10.8	14.8	0.383	0.287	0.254	21.6	237	4.20	0.55	0.37	0.00
21	2-Methyl-2-bropanol	27.1	38.0	0.389	0.767	0.234	22.0	247	5.20	0.41	1.01	0.68
22	Dimethylulfoxide	19.3	29.8	0.444	0.941	0.283	20.1	362	3.20	1.00	0.76	0.00
23	Cyanomethane	19.3	14.1	0.460	0.924	0.211	11.1	160	5.20	0.75	0.31	0.19
24	Nitromethane	20.5	2.7	0.481	0.926	0.233	12.5	65	5.10	0.85	0.134	0.22
25	2-Propanol	33.5	36.0	0.546	0.852	0.230	17.7	236	8.70	0.48	0.95	0.76
26	Butanol	36.8	29.0	0.586	0.842	0.242	22.1	231	10.30	0.47	0.88	0.79
27	Ethanoic acid	52.9	20.0	0.648	0.631	0.227	13.0	139	14.60	0.64	0.495	1.12
28	Ethanol	37.1	32.0	0.654	0.886	0.221	12.9	235	11.60	0.54	0.77	0.83
29	Methanol	41.3	30.0	0.762	0.913	0.203	8.2	218	14.90	0.60	0.62	0.93
30	Deuterium oxide[c]	54.8	33.0	0.991	0.963	0.206	3.7	156	21.80	1.09	0.18	1.17

[a] Except for *AN* and *DN*, parameters are for heptane.
[b] Parameters for trichloromethane.
[c] Parameters for water.
[d] Scaled by a factor of 0.01.

One of the questions that can be, and in fact has been, raised is the exact nature of the information contained in each of these descriptors, and whether or not there is any 'cross-information content,' that is, for example, do any of the polarity scales contain acidity or basicity information, and presenting that information in fundamental chemical meaning. The relevance of these scales is shown in a recent paper, where several of these solvent scales were applied to polymer properties.[50] The unique nature of the LSER methodology as contained within the TLSER permits this issue to be addressed, and to provide fundamental descriptions in terms of readily understandable parameters and descriptions. This study examines the Reichardt E^{T}_{N}, the Swain acidity (*E*) and basicity (*B*) parameters, the Koppel Palm polarity (*Y*) and polarizability (*P*) parameters, the molar refraction (*MR*) and the Gutmann acceptor number (*AN*) and donor number (*DN*).

As has been noted in previous TLSER work, the MNDO algorithm has been found to provide better regressions and is more in line with chemical intuition than has currently been found with either AM1 or PM3.[37,51] This has been attributed to the nature of the error functions within MNDO and AM1/PM3. Although larger, the error within MNDO is generally systematic and, as such, can be factored out into the intercept in a LFER regression. The error in AM1 or PM3, on the other hand, is random and, although much smaller, cannot be removed through regression analysis. Therefore, although AM1 and PM3 are much better for 'absolute' computations of properties, MNDO should still preferred when performing LFER or QSAR, and is used in the generation of all descriptors in this paper.

Although the TLSER parameters are computed for gas-phase, isolated molecules, it has been shown that these computed parameters are very applicable to correlations involving multiple solutes. Although the LSER methodology has required different parameters for a compound when behaving as a solute vs a solvent (for associated solvent), this can be attributed to the experimental nature of the LSER parameters. In the case of computed descriptors, of course, there is no distinction between solute and solvent, the parameters being computed in the gas phase only. Where the perturbation (solute vs solvent) is an effect of the solvation, the gas-phase generated parameters should be valid, and there-

Table 4. TLSER regressions for solvent scales[a] log (property) = $aV_{mc} + b\pi_I + c\epsilon_\beta + d_{q_-} + e\varepsilon_\alpha + f_{q_+} + g$

Parameter	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*N*	*R*	*F*	SD	R_{cv}^2
E_T^N	−0.3662	−4.0	n/a	0.3202	n/a	1.5164	0.8585	29	0.961	72	0.07468	0.8783
	−7.35	−3.30		3.28		6.26						
	1.3	1.3		1.2		1.4						
DN	n/a	n/a	n/a	41.356	−307.36	73.42	47.760	27	0.917	40	5.534	0.7875
				5.57	−4.91	4.21						
				1.1	1.1	1.2						
AN	n/a	−651.16	n/a	n/a	326.68	174.41	20.580	25	0.972	120	3.846	0.9112
		−6.95			4.85	14.47						
		2.8			2.6	1.2						
Y	−0.3337	n/a	n/a	0.962	3.070	1.369	0.1027	25	0.914	25	0.1144	0.738
	−4.15			4.75	2.24	3.01						
	1.3			1.3	1.2	1.7						
MR	22.659	181.85	n/a	n/a	n/a	n/a	−15.584	28	0.975	241	2.036	0.9253
	16.56	5.80										
	1.2	1.2										
P	n/a	2.0718	0.9021	n/a	n/a	n/a	−0.0868	28	0.925	74	0.126	0.8200
		8.80	3.80									
		1.2	1.2									
E	−11.844	n/a	n/a	n/a	−102.69	36.374	27.722	18	0.956	49	1.896	0.8253
	−6.02				−3.50	4.63						
	1.1				1.4	1.5						
B	n/a	N/A	2638	547.48	−2591.60	N/A	66.10	25	0.928	44	45.54	0.7993
			2.58	8.93	−4.84							
			1.1	1.1	1.1							

[a] In each case, the values given are, from top to bottom are coefficient value, *t*-score and VIF.

fore the same parameters for solute and solvent should be valid. Where the perturbation causes a difference in solvation, then the TLSER model is probably not valid and should not be used, nor should any gas-phase generated descriptors.

METHODOLOGY

The data for the computed solvent scales were taken from the original sources[52–56] as collated in a previous TLSER paper.[47] All geometries were optimized using the MNDO algorithm of MOPAC v6.0.[57,58] All TLSER parameters were extracted using the in-house developed MADCAP.[59] Structures were entered and optimized geometries examined using both PCMODEL and MMADS.[59–61] All statistical analysis was performed using MINITAB.[62] Cross-validated *R*-squared (R_{cv}^2) was determined using CODESSA.

All correlations reported adhere to three conditions. First, the cross-correlation of descriptors is minimized, as determined by the variance inflation factor (*VIF*).[63] A value of 1.0 is indicative of no correlation among the independent variables. A value of under 5.0 is considered acceptable, and over 10.0 indicates an unstable regression where the cross-correlation could yield statistically meaningless correlations. Second, all independent variables in the regression have a significance of at least 95% (based on Student's two-tailed *t*-score). Third, outliers are removed only if the normalized residual (observed minus predicted values) is greater than two standard deviations.

RESULTS AND DISCUSSION

The properties correlated using the TLSER are given in Table 1. Table 2 gives the TLSER parameters for each of the 30 compounds studied in this paper and Table 3 lists the solvent scale values for each of the compounds. Table 4 lists the resulting TLSER regressions for each of the eight solvent scales examined in this study. The parameters may be nominally divided into polarity/polarizability (E^T_N, *Y*, *P*, *MR*), acidity (*AN*, *E*) and basicity (*DN*, *B*). However, as will be described in the ensuing sections, there is significant 'cross content' of information in most of the solvent scales.

In addition to the coefficients for each of the parameters used in each relevant regression, Table 4 also lists the necessary statistical parameters for each parameter and/or regression. For each parameter, the coefficient, *t*-score and *VIF* are presented for each statistically significant parameter (based on the two-tailed *t*-score at the 95% level). For each regression, the number of data points (*N*), the value of the *F* statistic (*F*), the standard deviation for the regression (SD), and the cross-validated R^2 (R_{cv}^2) is provided.

Table 5. Observed, predicted and residual values for E^{T}_{N}

No.	Observed	Predicted	Residual
1	0.006	0.051	−0.045
2	0.009	0.030	−0.021
3	0.012	−0.038	0.050
4	0.043	0.073	−0.030
5	0.052	0.108	−0.056
6	0.071	−0.016	0.087
7	0.099	0.118	−0.019
8	0.111	0.168	−0.057
9	0.117	0.243	−0.126
10	0.188	0.143	0.045
11	0.207	0.291	−0.084
12	0.231	0.232	−0.001
13	0.256	0.298	−0.042
14	0.269	0.351	−0.082
15	0.302	0.280	0.022
16	0.324	0.189	0.135
17	0.327	0.283	0.044
19	0.355	0.352	0.003
20	0.383	0.272	0.111
21	0.389	0.504	−0.115
22	0.444	0.476	−0.032
23	0.460	0.380	0.080
24	0.481	0.424	0.057
25	0.546	0.578	−0.032
26	0.586	0.510	0.076
27	0.648	0.722	−0.074
28	0.654	0.660	−0.006
29	0.762	0.754	0.008
30	0.991	0.888	0.103

Table 6. Observed, predicted and residual values for *MR*

No.	Observed	Predicted	Residual
1	27.2	27.7	−0.5
2	34.5	29.6	4.9
3	34.5	33.7	0.8
5	26.4	32.9	−6.5
6	41.0	40.0	1.0
7	31.1	29.5	1.6
8	26.3	25.5	0.8
9	22.4	23.0	−0.6
10	31.0	29.5	1.5
11	20.0	20.9	−0.9
13	21.3	21.8	−0.5
14	16.4	17.0	−0.6
15	24.2	24.2	−0.0
16	32.8	31.2	1.6
17	21.0	21.2	−0.2
18	31.4	30.2	1.2
19	16.2	16.7	−0.5
20	21.6	22.9	−1.3
21	22.0	22.4	−0.4
22	20.1	19.8	0.3
23	11.1	11.7	−0.6
24	12.5	15.0	−2.5
25	17.7	18.1	−0.4
26	22.1	22.6	−0.5
27	13.0	13.7	−0.7
28	12.9	13.5	−0.6
29	8.2	8.3	−0.1
30	3.7	−0.1	3.8

Table 7. Observed, predicted and residual values for Koppel Palm *Y*

No.	Observed	Predicted	Residual
1	0.254	0.213	0.041
2	0.235	0.175	0.060
3	0.235	0.121	0.114
4	0.321	0.455	−0.134
5	0.292	0.549	−0.257
6	0.407	0.366	0.041
7	0.315	0.481	−0.166
8	0.300	0.494	−0.194
9	0.526	0.587	−0.061
10	0.606	0.535	0.071
11	0.681	0.635	0.046
12	0.667	0.614	0.053
13	0.559	0.647	−0.088
14	0.729	0.675	0.054
15	0.790	0.720	0.070
16	0.918	0.770	0.148
17	0.757	0.646	0.111
19	0.868	0.722	0.146
21	0.767	0.794	−0.027
24	0.926	0.893	0.033
25	0.852	0.861	−0.009
26	0.842	0.804	0.038
28	0.886	0.918	−0.032
29	0.913	0.975	−0.062
30	0.963	0.959	0.004

Table 8. Observed, predicted and residual values for Koppel Palm *P*

No.	Observed	Predicted	Residual
1	0.256	0.247	0.009
2	0.236	0.232	0.004
3	0.236	0.236	0.000
4	0.243	0.258	−0.015
5	0.274	0.258	0.016
6	0.242	0.246	−0.004
7	0.293	0.301	−0.008
8	0.295	0.299	−0.004
9	0.217	0.242	−0.025
10	0.306	0.305	0.001
11	0.247	0.248	−0.001
12	0.231	0.250	−0.019
13	0.252	0.249	0.003
14	0.256	0.236	0.020
15	0.299	0.296	0.003
16	0.322	0.314	0.008
17	0.266	0.239	0.027
18	0.308	0.310	−0.002
20	0.254	0.256	−0.002
21	0.234	0.236	−0.002
22	0.283	0.262	0.021
23	0.211	0.213	−0.002
24	0.233	0.257	−0.024
25	0.230	0.233	−0.003
26	0.242	0.233	0.009
27	0.227	0.228	−0.001
28	0.221	0.224	−0.003
29	0.203	0.210	−0.007

Table 9. Observed, predicted and residual Values for Gutmann *AN*

No.	Observed	Predicted	Residual
2	0.0	0.1	−4.1
3	0.0	3.1	−3.1
4	1.4	6.7	−5.3
5	8.6	5.5	3.1
8	8.2	9.4	−1.2
9	3.9	4.6	−0.7
11	8.0	5.9	2.1
13	23.1	23.7	−0.6
14	20.4	20.7	−0.3
15	14.2	15.2	−1.0
16	14.8	10.5	4.3
17	16.7	19.4	−2.7
18	15.5	9.5	6.0
19	12.5	16.9	−4.4
20	10.8	6.2	4.6
21	27.1	35.0	−7.9
22	19.3	17.9	1.4
23	19.3	16.2	3.1
24	20.5	17.5	3.0
25	33.5	36.4	−2.9
26	36.8	36.3	0.5
27	52.9	51.4	1.5
28	37.1	38.5	−1.4
29	41.3	41.8	−0.5
30	54.8	48.4	6.4

Table 10. Observed, predicted and residual values for Swain *E*

No.	Observed	Predicted	Residual
2	0.06	−1.20	1.26
7	1.30	−0.21	1.51
9	2.10	1.95	0.15
13	3.28	2.99	0.29
14	2.70	4.37	−1.67
17	3.00	1.82	1.18
19	2.10	3.38	−1.28
20	4.20	3.61	0.59
21	5.20	8.74	−3.54
22	3.20	3.18	0.02
23	5.20	6.46	−1.26
24	5.10	5.25	−0.15
25	8.70	10.85	−2.15
26	10.30	8.65	1.65
27	14.60	11.95	2.65
28	11.60	13.15	−1.55
29	14.90	15.56	−0.66
30	21.80	18.83	2.97

Nominal polarity/polarizability parameters

Rows 1, 4, 5 and 6 of Table 4 show the regressions of the solvent scales that have been attributed to solvent polarity or polarizability. Tables 5, 6, 7 and 8 list the observed, predicted and residual values for each of the compounds (non-outliers) used in the developing each of the regressions. For the E^{T}_{N}, chlorobenzene was identified as an outlier and removed from the final regression. Tetrachloromethane, bis(2-methoxyethyl) ether and cyanobenzene were likewise identified as outliers and removed. For molar refractivity, triethylamine and bis(2-methoxyethyl) ether were removed as outliers. For *P*, propanone and deuterium oxide were removed.

Although there are similarities in each of these scales, based on the TLSER, there are also significant differences. E^{T}_{N} and *MR* include both the molecular volume and the polarizability index, while *Y* and *P* use only the V_{mc} and *P* uses only π_{I}. As has been demonstrated by Abraham, and shown previously for the TLSER, volume and polarizibility taken together are indicative of dispersion effects.

E^{T}_{N} also has a significant contribution from both the electrostatic acidity and basicity. This is indicative of dipolar effects and a permanent dipolarity contribution. That q_{+} is more than twice as significant (based on the Student's *t*-score, as provided in Table 4) as q_{-} also indicates that there are additional acidity factors being included in this solvent scale. This indicates that although E^{T}_{N} contains not only a measure of the polarity of a solvent, it also contains information concerning polarizability, in addition to acidity. Therefore, these additional effects must be taken into account when E^{T}_{N} is used in explaining solvent-based phenomena.

Table 11. Observed, predicted and residual values for Gutmann *DN*

No.	Observed	Predicted	Residual
1	0.0	3.6	−3.6
2	0.0	4.4	−4.4
3	0.0	4.3	−4.3
6	19.0	17.9	1.1
7	0.1	2.3	−2.2
8	0.1	1.0	−0.9
9	19.2	17.7	1.5
10	3.3	2.9	0.4
11	20.0	17.6	2.4
12	20.0	25.8	−5.8
13	4.0	2.0	2.0
14	1.0	4.0	−3.0
16	4.4	10.4	−6.0
17	0.0	4.5	−4.5
18	11.9	0.1	11.8
19	17.0	8.5	8.5
20	14.8	18.1	−3.3
21	38.0	29.5	8.5
22	29.8	27.9	1.9
23	14.1	4.2	9.9
24	2.7	9.4	−6.7
25	36.0	29.5	6.5
26	29.0	29.9	−0.9
27	20.0	26.6	−6.6
28	32.0	30.4	1.6
29	30.0	31.5	−1.5
30	33.0	35.2	−2.2

Table 12. Observed, predicted and residual values for Swain *B*

No.	Observed	Predicted	Residual
1	0	30	−30
2	0	32	−32
3	0	32	−32
6	285	231	54
7	58	68	−10
8	48	46	2
9	280	236	44
10	38	55	−17
11	287	227	60
12	238	230	8
13	14	−46	60
14	23	12	11
16	67	139	−72
17	40	29	11
19	224	141	83
20	237	226	11
21	247	220	27
22	362	399	−37
24	65	121	−56
25	236	217	19
26	231	217	14
27	139	168	−29
28	235	222	13
29	218	229	−11
30	156	249	−93

For *Y* likewise, there is an electrostatic dipolarity contribution indicated by the two positive contributions from q_+ and q_-. Like the Reichardt parameter, there is an additional acidity contribution, but the acidity component in *Y* appears to be primarily covalent and not electrostatic, as is evidenced by the contribution of ε_β. Neither *P* or *MR* has an electrostatic dipolar or an acidity contribution. *MR* appears to be almost entirely dispersion based, which is consistent with the previous literature. The primary contribution in *P* appears to be the polarizability, with a small contribution from basicity. The former is consistent with the original intent of the Koppel Palm descriptor.

Nominal acidity scales

The third and seventh rows of Table 4 show the regressions for the Gutmann *AN* and Swain *E*, respectively. Tables 9 and 10 list the observed, predicted and residual values for these two scales. In both of these regressions, the electrostatic acidity is the primary contribution. For the *AN*, the covalent acidity also plays an important and contributory role. In the *E*, however, increased covalent acidity reduces the Swain *E* value. The negative contributions from the π_I (*AN*) and V_{mc} (*E*) indicate that delocalization and dispersion reduce the acidity. Therefore, while the *AN* provides a complete description of the acidity, while the Swain *E* is appears to be measuring only the electrostatic acidity component.

Nominal basicity scales

Rows 2 and 8 of Table 4 show the regression for the two basicity scales, Gutmann *DN* and Swain *B*, respectively. Tables 11 and 12 lists the observed, predicted and residual values for these two basicity parameters. Only triethylamine was identified as an outlier and removed from the *DN* regression. Triethylamine, pyridine, carbon tetrachloride, cyanobenzene and acrylonitrile were outliers for *B*. For *DN*, the primary contribution appears to be from a dipolarity contribution, as is evidenced by the fairly equal contribution of the (both positive) q_- and q_+. The primary basicity appears to be covalent, and arises from the negative coefficient of ε_α. A negative ε_α indicates that lower acidity increases the *DN* value, which is consistent with the concept of *DN*. *B*, on the other hand, has no positive acidity coefficient, i.e. there is not dipolarity contribution, and the entire contribution comes from covalent or electrostatic basicity, or a negative acidity contribution. Therefore, although *B* appears to be a completely basicity-based parameter, *DN* once again appears to contain not only a basicity (or negative acidity) contribution, but also electrostatic dipolarity.

CONCLUSION

Katritzky *et al.*[48] have developed similar regressions to each of these solvent scales using a different approach. A direct comparison of the Katritzky method with the TLSER is not possible because of the different data sets used to develop the regression and the use of AM1 vs MNDO.

The TLSER has been shown in this study to correlate with several commonly used solvent polarity, acidity and basicity scales. Further, the TLSER is provides not only an equation useful for the prediction of solvent parameters, but also a mechanism for probing the fundamental properties which constitute the various solvent scales. This latter capability is possible because the TLSER parameters have a specific theoretical underpinning and provide specific meaning at the molecular level.

REFERENCES

1. C. Hansch and T. J. Fujita, *J. Am. Chem. Soc.* **86**, 1616 (1964).
2. C. Hansch, *Acc. Chem. Res.* **2**, 232 (1969).
3. C. Hansch and A. Leo, *Correlation Analysis in Chemistry and Biology*. Wiley–Interscience, New York (1979).
4. I. Koppel and V. Palm, in Shorter, J. and N. Chapman. *Advances in Linear Free Energy Relationships*, edited by I. Koppel and V. Palm, p. 203. Plenum Press, London (1972).
5. C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry*. VCH, New York (1988).
6. W. L. Jorgensen and C. J. Ravimohan, *J. Chem. Phys.* **83**, 3050 (1985).

7. C. J. Cramer and D. G. Truhlar, *Science* **256**, 213 (1992).
8. A. T. Balaban, *Chem. Phys. Lett.* **89**, 399 (1982).
9. O. Exner, *Correlation Analysis of Chemical Data*. Plenum Press, New York (1988).
10. W. J. D. Dunn, *Toxicol. Lett.* **43**, 277 (1988).
11. G. R. Famini, R. J. Kassel, J. W. King and L. Y. Wilson, *Quant. Struct. -Act. Relat.* **10**, 344 (1991).
12. G. R. Famini, W. P. Ashman, A. P. Miskiewicz and L. Y. Wilson, *Quant. Struct.-Act. Relat.* **11**, 162 (1992).
13. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure Activity Analysis*. Research Studies Press, Letchworth (1986).
14. A. R. Katritzky, M. Karelson and V. S. Lovanov, *J. Chem. Inf. Comput. Sci.* **69**, 245 (1997).
15. P. Politzer and J. Murray, *Quantitative Treatments of Solute/ Solvent Interactions*. Elsevier, Amsterdam (1994).
16. P. Politzer, J. Murray, P. Lane and T. Brink, *J. Phys. Chem.* **97**, 5144 (1992).
17. H. Koneman, *Toxicity* **19**, 209 (1981).
18. D. F. V. Lewis, *Prog. Drug Metabolism* **12**, 205 (1990).
19. D. F. V. Lewis, in , edited by D. F. V. Lewis, in Reviews of computational chemistry, K. Lipkowitz and D. Boyd, eds. Vol. **3**, p. 173 (1992).
20. S. C. Basak and G. D. Grunwald, in Proceedings of the XVI International Cancer Congress, p. 413 (1995).
21. M. J. Kamlet, R. W. Taft and J.-L. M. Abboud, *J. Am. Chem. Soc.* **91**, 8325 (1977).
22. M. J. Kamlet and R. W. Taft, *Prog. Org. Chem.* **48**, 2877 (1983).
23. M. J. Kamlet, R. W. Taft, G. R. Famini and R. M. Doherty, *Acta Chem. Scand.* **41**, 589 (1987).
24. R. W. Taft, M. H. Abraham, G. R. Famini, R. M. Doherty, J. L. Abboud, and M. J. Kamlet, *J. Pharm. Sci.* **74**, 807 (1985).
25. M. H. Abraham, G. S. Whiting, R. Fuchs and E. J. Chambers, *J. Chem. Soc., Perkin Trans. 2* 291 (1990).
26. M. H. Abraham, R. Kumarsingh, J. E. Cometto-Muniz, and W. S. Cain, *Arch. Toxicol.* **72**, 227 (1998).
27. J. Debord, T. Dantoine, J. C. Bollinger, M. H. Abraham, B. Verneuil and L. Merle, *Chem. Biol. Interact.* **113**, 105 (1998).
28. M. H. Abraham, K. Takacs-Novak and R. C. Mitchell, *J. Pharm. Sci.* **86**, 310 (1997).
29. M. G. Ford and D. J. Livingstone, *Quant. Struct.–Act. Relat.* **9**, 107 (1990).
30. T. F. Wolszyn and P. C. Jurs, *Anal. Chem.* **64**, 3059 (1992).
31. A. R. Katritzky and E. V. Gordeeva, *J. Chem. Inf. Comput. Sci.* **33**, 835 (1993).
32. D. R. Passino-Reader, J. P. Hickey and L. M. Ogilvie, *Bull. Environ. Contam. Toxicol.* **59**, 834 (1997).
33. G. R. Famini, *Using Theoretical Descriptors in Quantitative Structure Activity Relationships V*, Vol. CRDEC-TR-095. US, Army Edgewood Research Development and Engineering Center, Aberdeen Proving Ground, MD (1989).
34. G. R. Famini and L. Y. Wilson, *J. Phys. Org. Chem.* **6**, 539 (1993).
35. G. R. Famini, C. E. Penski and L. Y. Wilson, *J. Phys. Org. Chem.* **11**, 1992 (1992).
36. G. R. Famini and L. Y. Wilson, in *Using Theoretical Descriptors in LSER*, edited by G. R. Famini and L. Y. Wilson, Vol. **1**, p. 213. Elsevier, Amsterdam (1995).
37. G. R. Famini and L. Y. Wilson, *Chemosphere* (1997).
38. L. Y. Wilson and G. R. Famini, *J. Med. Chem.* **34**, 1668 (1991).
39. G. R. Famini and L. Y. Wilson, *Org. React.* **29**, 117 (1995).
40. A. H. Headley, S. D. Starnes, L. Y. Wilson and G. R. Famini, *J. Org. Chem.* **59**, 8040 (1994).
41. A. H. Lowrey and G. R. Famini, *Struct. Chem.* **6**, 357 (1995).
42. A. H. Lowrey, G. R. Famini, V. Loumbev, L. Y. Wilson and J. M. Tosk, *Pigm. Cell Res.* **10**, 251 (1997).
43. J. S. Murray, S. Ranganathan and P. Politzer, *J. Org. Chem.* **56**, 3734 (1991).
44. J. S. Murray and P. Politzer, *J. Org. Chem.* **56**, 6715 (1991).
45. J. Murray and P. Politzer, *J. Chem. Res.* 110 (1992).
46. J. S. Murray, P. Politzer and G. R. Famini, *J. Mol. Struct. (Theochem)* **454**, 299 (1998).
47. J. B. F. N. Engberts, G. R. Famini, A. Perjessy and L. Y. Wilson, *J. Phys. Org. Chem.* **11**, 261 (1998).
48. A. R. Katritzky, T. Tamm, Y. Wang, S. Sild and M. Karelson, *J. Chem. Inf. Comput. Sci. submitted*, for publication.
49. A. R. Katritzky, V. S. Lobanov and M. Karelson, CODESSA Reference Manual Version 2.0. University of Florida, Grimesville (1994).
50. S. Spange, A. Reuter and W. Linert, *J. Polym. Sci. Part A, Polym. Chem.* **36**, 1945 (1998).
51. W. H. Donovan and G. R. Famini, *J. Chem. Soc. Perkin Trans. 2* 83 (1996).
52. V. Gutmann, *Electrochim. Acta* **21**, 661 (1976).
53. Y. Marcus, *J. Solution Chem.* **13**, 599 (1984).
54. Y. Marcus, *Chem. Soc. Rev.* **23**, 409 (1993).
55. V. Palm and N. Palm, *Org. React.* **28**, 2319 (1993).
56. C. Reicardt, *Chem. Rev.* **94**, 2319 (1994).
57. M. J. K. Dewar and W. Thiel, *J. Am. Chem. Soc.* **99**, 4899 (1977).
58. J. J. P. Stewart, MOPAC Manual, Vol. FJSRL-TR-88-007. J. Frank Seiler Research Laboratory, Colorado Springs, CO (1988).
59. G. R. Famini and J. A. Fager, in *MOPAC Automated Data Collection and Assembly Program*, edited by G. R. Famini and J. A. Fager, US Army Edgewood Research, Development and Engineering Center, Aberdeen Proving Ground, MD (1995).
60. PCModel. Serena Software, Bloomington, IN.
61. J. M. Leonard and G. R. Famini, A User's Guide to MMADS, Vol. CRDEC-TR-030. US, Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD (1989).
62. Minitab. State College, PA Minitab, Inc.
63. D. A. Belesley, E. Kuh and R. E. Welsh, *Regression Diagnostics*. Wiley, New York (1980).

