# Using Theoretical Descriptors in Quantitative Structure–Activity Relationships: Gas Phase Acidity

**George R. Famini,**[a] **Benjamin C. Marquez**[b] **and Leland Y. Wilson***[,b]
[a] *U.S. Army Edgewood Research, Development and Engineering Center, Aberdeen Proving Ground, MD 21010, USA*
[b] *Department of Chemistry, La Sierra University, Riverside, CA 92515, USA*

The application of computational techniques to biology, chemistry and physics is growing rapidly. Quantitative structure–activity relationships (QSAR) have been used widely to relate biological activities as well as physicochemical properties to molecular structure features. A difficulty in this approach has been nonuniformity of parameter sets resulting in the inability to examine contributions across properties and data sets. Linear solvation energy relationships (LSER) developed by Kamlet and Taft successfully utilize a single set of parameters to correlate a wide range of biological, chemical and physical properties. We have replaced the empirical LSER solvatochromic parameters with theoretically determined parameters to permit greater ease in *a priori* property prediction. These TLSER descriptors have given good correlations and interpretations for some biological activities and physicochemical properties. This paper discusses the application of these descriptors to the gas phase acidity of some carboxylic acids, alcohols, silanols, anilines, hydrocarbons, and oximes. Good correlations and physical interpretations result.

Quantitative structure–activity (property) relationships (QSAR, QSPR) have been used to correlate molecular structural features of compounds with their known biological, chemical and physical properties. QSAR assumes that there is a quantitative relation between microscopic (molecular structural) features and a macroscopic (empirical) property of a compound. Originally, activity and QSAR referred to a biological property; however, the QSAR concepts apply equally well to physicochemical properties. Once a relation is found for a particular property, it can be used to predict that property for any compound from its molecular structure.[1] One such equation is based on the *linear free energy relationship* (LFER). Burkhardt[2] and Hammett[3] reviewed the existence of LFERs in 1935; in 1937 Hammett[4] proposed the equation that bears his name. A recent (1988) survey of LFER and a clear discussion of the background for its use is given by Exner.[5]

*Linear Solvation Energy Relationships.*—An enormous number of descriptors has been used by researchers to increase the ability to correlate biological, chemical and physical properties. One of the most successful sets has been used in the correlations of Hansch[6] and Kamlet, Taft and co-workers[7] who extended the LFER of earlier workers[8] to involve solute–solvent interactions.[9] This *linear solvation energy relationship* (LSER) has the general form shown in eqn. (1).

$$\text{property} = \text{bulk/cavity term} + \text{dipolarity/polarizability} + \text{hydrogen bonding term(s)} + \text{constant} \quad (1)$$

The property is often the logarithm of a measured quantity (involving solute–solvent interactions) which, in turn, can be related to a free energy consistent with the LFER concept. Kamlet, Taft and co-workers employed the empirically based solvatochromic (LSER) descriptor set for the terms in eqn. (1). For a given property of solutes in a given solvent, the bulk term uses the solute intrinsic volume, $V_I$, the dipolarity terms use the solute dipolarity parameter, $\pi^*$, and the solute polarizability correction, $\delta$, while the hydrogen bonding terms employ a solute acidity descriptor, $\alpha$, and a solute basicity descriptor, $\beta$. Early work used the molar volume, $V_m$, while $V_I$ is computed. For a property of solvents with a given solute, the Hildebrand solubility parameters, $\delta_H{}^2$, replaces the volume in the bulk term while the other parameters, now pertaining to the solvent, are retained. Subscripts 1 and 2 refer to solvent and solute respectively. Usually not all the terms in eqn. (1) are statistically significant.

These LSER terms can be interpreted in microscopic (energetic, bonding) and macroscopic (thermodynamic) terms. The bulk term is a measure of the energy needed to overcome cohesive solvent–solvent molecule interactions (endoergic) to form a cavity for the solute molecule. The dipolarity polarizability terms are measures of the energies of solute–solvent dipole and induced dipole interactions (exoergic) which contribute to solution formation. Hydrogen bonding terms measure the energy of interaction (exoergic) when a solute–solvent complex is formed. Respectively, hydrogen bond acceptor basicity (HBAB) and hydrogen bond donor acidity (HBDA) refer to accepting/donating a proton from/to a neighbour molecule in keeping with the Brønsted–Lowry acid–base definitions.

Thermodynamic interpretation can be inferred from work by Abraham and co-workers[10] who correlated thermodynamic quantities for the process, X(water) $\longrightarrow$ X(hexadecane), with the LSER descriptors. The volume term (bulk) is related to the difference in energy needed to create a solute molecule sized cavity in the two solvents; it is endoergic in each solvent. If the energy is greater in the water it will make the standard enthalpy change for the process more exothermic. The cavity formation seems to involve general dispersive forces also; these will be more exoergic and exothermic in a nonpolar solvent than in water, again contributing to a more exothermic overall change. The dipolarity and polarizability terms seem not to be as easily interpreted thermodynamically; however, greater dipolarity implies a greater tendency to form solute–water dipole–dipole interactions which are expected to be exoergic. The hydrogen bonding terms involve the difference between the exothermic solute–water interaction and the much less exothermic solute–hexadecane interaction resulting in an overall endothermic enthalpy of transfer. However, the entropy change from the formation of solute–water bonds will be less than that for solute–hexadecane bonds resulting in an overall positive entropy change.

A strongpoint of these solvatochromic (LSER) descriptors is their very successful correlation of more than 200 biological, chemical and physical properties involving solute–solvent interactions for a large number of compounds.[11] The coefficients of the descriptors in the correlation equation can also provide insight into the nature of the solute–solvent interactions as typified by the discussion in the previous paragraphs.

However, the LSER descriptors are somewhat limited in their ability to make *a priori* predictions because they are empirical. Although there are tables of LSER parameters and predictive relations to help in their estimation, LSER values for complex molecules are not as easily found. Hickey and Passino-Reader have provided 'rules of thumb' for LSER parameter estimation.[12]

Closely related to the LSER approach is a four-parameter correlation recently reported by Gajewski[13] who indicates good success for solvent rate effects using the KOPMH (Kirkwood, Onsager, Parker, Marcus, Hildebrand) equation. The four solvent parameters are the Kirkwood–Onsager function, $(\varepsilon - 1)/(2\varepsilon + 1)$,[14] the Hildebrand solvent cohesive energy density, $(\Delta H_{vap} - RT)/V_m$,[15] (this is the $\delta_H{}^2$ mentioned earlier) and anion and cation relative stabilization parameters, $\alpha'$ and $\beta'$, respectively. For chloride and potassium ions, values for these last two parameters are derived from Parker's and Marcus'[16] compilations of relative free energy of solvation in several solvents. For other solvents not covered by Parker and Marcus, $\alpha'$ and $\beta'$ are estimated from the Kamlet–Taft (LSER) $\alpha$ and $\beta$.

Attempts to correlate computationally derived structural and electronic descriptors with the solvatochromic parameters have met with moderate degrees of success by Lewis.[17] Recently Politzer has related the molecular electrostatic potential to the LSER dipolarity/polarizability index[18] and solute hydrogen bonding descriptors.[19]

*Theoretical Linear Solvation Energy Relationships.*—In the past, theoretical chemistry has been used to provide descriptors for QSAR. Representative work is described by Loew and coworkers,[20] Pedersen,[21] and Chastrette and colleagues[22] while Lewis[23] gives a more recent summary of molecular orbital calculations applied to QSAR (MO–QSAR) for a variety of activities. Examples of theoretical descriptors include the molecular transform of Kier and Hall.[24] Ford and Livingstone[25] point out advantages of computationally derived descriptors over extra-thermodynamically derived descriptors such as pi and sigma. They are not restricted to closely related compounds as is often the case with group theoretical, topological and other variables. They describe clearly defined molecular properties making the interpretation of QSAR equations more straightforward. Furthermore, their values are easily obtained; no laboratory measurements are needed thus saving time, space, materials, equipment and alleviating safety (toxicity) and disposal concerns.

Based on the LSER philosophy and general structure a new, theoretical set of parameters for correlating a wide variety of properties has been developed.[26] These theoretical linear solvation energy relationship (TLSER) descriptors have shown good correlations and physical interpretations for the following: five nonspecific toxicities;[27] activities of some local anaesthetics and the molecular transform;[28] opiate receptor activity of some fentanyl-like compounds;[29] and six physicochemical properties: charcoal absorption; HPLC retention index; octanol–water partion coefficient; phosphonothiolate hydrolysis rate constant; aqueous acid equilibrium constant; and electronic absorption of some ylides.[30] These TLSER parameters are determined solely from computational methods thus permitting *a priori* prediction of properties. The TLSER descriptors were developed to correlate closely with the LSER descriptors; to give equations with correlation coefficients, *R*, and standard deviations, SD, close to those for LSER; and to be as widely applicable to solute–solvent interactions as the LSER set. Table 1 gives a summary of TLSER descriptors as used in this paper.

The TLSER bulk/steric term is described by the molecular van der Waals volume, $V_{mc}$, in units of 100 cubic angstroms. The dipolarity/polarizability term uses the polarizability index, $\pi_I$, obtained by dividing the polarizability volume by the molecular volume to produce a unitless, size independent quantity which indicates the ease with which the electron cloud may be moved or polarized. For example, aromatics and chlorine rank high while alkanes and fluorine rank low on the scale.

The hydrogen bond acceptor basicity (HBAB) is composed of covalent, $\varepsilon_b$, and electrostatic, $q_-$, basicity terms. Analogously, the hydrogen bond donor acidity (HBDA) is made up of covalent, $\varepsilon_a$, and electrostatic, $q_+$, acidity terms. The covalent HBAB parameter, $\varepsilon_b$, is the magnitude of the difference between the energy of the highest occupied molecular orbital (HOMO) of the solute and the lowest unoccupied molecular orbital (LUMO) of water. The result is divided by 100 for convenience in presentation and comparison of coefficients; the units are in hectoelectron volts (heV). Analogously, the covalent HBDA parameter, $\varepsilon_a$, is the magnitude of the difference between the energies of the LUMO of the solute and the HOMO of water again scaled like the covalent HBAB with the same units. The water energies are included for aesthetic reasons; the smaller these differences the greater is the ability to form a hydrogen bond with water. The electrostatic contribution to the HBAB is the magnitude of the largest negative formal charge, $q_-$, on an atom; units are atomic charge units (acu). The corresponding HBDA descriptor is the formal charge, $q_+$, on the most positively charged H atom (again in acu). The values of $q_-$ and $q_+$ are obtained by Mulliken population analysis.

Previous papers used $\varepsilon_b$ and $\varepsilon_a$; however, increasing $\varepsilon_b$ and $\varepsilon_a$ mean decreasing basicity and acidity, respectively. Consequently, this paper, uses the transformation, $\varepsilon_{A \text{ or } B} = 0.300 - \varepsilon_{a \text{ or } b}$, so that $\varepsilon_A$ and $\varepsilon_B$ increase with increasing acidity or basicity; the 0.300 value was chosen to give values of similar size.

The generalized TSLER equation for solutes in a given solvent, is eqn. (2) where SSP represents a solute–solvent

$$\mathrm{SSP} = \mathrm{SSP}_0 + aV_{mc} + b\pi_I + c\varepsilon_B + dq_- + e\varepsilon_A + fq_+ \quad (2)$$

interaction property; this is generally taken as the logarithm of a measured quantity. For a given property and set of compounds, the coefficients ($\mathrm{SSP}_0$ through *f*) are determined using multilinear regression analysis to fit the data. In most cases not all terms are significant.

The theoretical descriptors are from calculations that, strictly speaking, apply to an isolated molecule and, thus the gas phase. With this in mind, and noting the aqueus acidity correlations mentioned earlier, Professor R. W. Taft[31] suggested that correlations of gas phase acidities and TLSER parameters be studied. Good TLSER correlations with properties involving solute–solvent interactions suggest that the gas phase parameters correlate well with properties effecting the solute–solvent system. In effect, the initial state (an isolated molecule) seems to influence strongly the final state (a solute–solvent complex). Related to this initial state influencing the property is the work of Siggel and co-workers[32] who have shown that the gas phase acidities of *p*-nitrophenols relative to phenol are influenced by the charge distribution on the initial state, the molecule, and not by the resonance delocalization on the final state, the anion. Thus, it seemed reasonable that gas phase acidity might correlate well with TLSER descriptors.

**Table 1** TLSER descriptors[a]

| Symbol | Name | Definition | Units | Meaning |
|---|---|---|---|---|
| $V_{mc}$ | molecular volume | molecular volume | 100 Å$^3$ | cavity/steric |
| $\pi_I$ | polarizability index | polarizability/$V_{mc}$ | none | polarizability |
| $\varepsilon_B$ | 'covalent' HB basicity | $0.30 - (\lvert(E(\mathrm{h}) - E(\mathrm{lw}))\rvert/100)$ | heV | HBAB |
| $q_-$ | 'electrostatic' HB basicity | maximum $\lvert(-)$ charge$\rvert$ on an atom | acu | HBAB |
| $\varepsilon_A$ | 'covalent' HB acidity | $0.30 - (\lvert(E(\mathrm{l}) - E(\mathrm{hw}))\rvert/100)$ | heV | HBDA |
| $q_+$ | 'electrostatic' HB acidity | maximum (+) charge on an H atom | acu | HBDA |

[a] Å = Angstrom; heV = hectoelectronvolt; acu = atomic charge unit; HB = hydrogen bond; $E$(l) = LUMO energy; $E$(h) = HOMO energy; $E$(lw) and $E$(hw) refer to the $E$(LUMO) and $E$(HOMO) for water, respectively; $\lvert\ \rvert$ indicate absolute magnitudes; HBAB = hydrogen bond acceptor basicity; HBDA = hydrogen bond donor acidity.

**Table 2** Correlations for individual sets (outliers kept)[a] $\Delta G_a{}^b = aV_{mc} + b\pi_I + c\varepsilon_B + dq_- + e\varepsilon_A + fq_+ + g$

| Coeff.<br>±<br>$t$-stat.<br>$P$(2-tail)<br>VIF | $a$ | $b$ | $c$ | $d$ | $e$ | $f$ | $g$ | $n$ | $R$ | SD | $F$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Acids | −17.6<br>8.1<br>2.16<br>0.042<br>1.29 | n/s[c] | n/s | 676<br>89<br>7.53<br>0.000<br>1.25 | −1654<br>359<br>4.60<br>0.000<br>1.20 | −1710<br>233<br>7.34<br>0.000<br>1.31 | 1751<br>81<br>21.6<br>0.000 | 27 | 0.937 | 8.4 | 39 |
| Alcohols | −31.9<br>4.7<br>6.76<br>0.000<br>1.09 | n/s | n/s | 753<br>81<br>9.35<br>0.000<br>1.51 | n/s | −2038<br>308<br>6.61<br>0.000<br>1.61 | 1686<br>76<br>22.3<br>0.000 | 21 | 0.975 | 7.2 | 107 |
| Silanols | n/s | n/s | n/s | 147<br>24<br>6.06<br>0.000<br>1.28 | n/s | −551<br>37<br>14.8<br>0.000<br>1.28 | 367<br>14<br>25.9<br>0.000 | 18 | 0.968 | 2.5 | 111 |
| Anilines | n/s | n/s | n/s | n/s | −998<br>187<br>5.35<br>0.000<br>1.00 | n/s | 536<br>34<br>16.0<br>0.000 | 13 | 0.850 | 2.1 | 29 |
| Oximes | n/s | n/s | n/s | n/s | −6011<br>1801<br>3.34<br>0.021<br>1.00 | n/s | 2541<br>328<br>7.76<br>0.001 | 7 | 0.831 | 39 | 11 |
| Hydrocarbons | n/s | n/s | −2011<br>166<br>12.1<br>0.000 | n/s | n/s | n/s | 318.4<br>23.3<br>13.7<br>0.000 | 16 | 0.955 | 9.6 | 147 |
| Small hydrocarbons | n/s | n/s | n/s | n/s | −5065<br>1425<br>3.56<br>0.021<br>1.00 | n/s | 2494<br>236<br>10.6<br>0.000 | 8 | 0.823 | 25 | 13 |

[a] Acids, no outliers; alcohols, $CH_3OCH_2CH_2OH$; silanols, $(CH_3)_3SiOH$; oximes, no outliers; hydrocarbons, thiophene; small hydrocarbons, ethene. [b] kJ mol$^{-1}$, $\Delta H_a$ used for oximes and hydrocarbons, activation energy for hydrocarbons, see discussion. [c] n/s, not significant at 0.95 level.

*Gas Phase Acidity*.—Gas phase acidity is defined in the following obvious manner.

$$RH(g) \longrightarrow H^{+1}(g) + R^{-1}(g); \Delta H_a$$

Quantitative data is given in terms of the standard enthalpy change, $\Delta H_a$, instead of the equilibrium constant. It is important to note that increasing enthalpy change means decreasing acidity. The $\Delta H_a$ values may be experimentally determined through high pressure mass spectrometry (HPMS), ion cyclotron resonance (ICR) and flowing afterglow (FA) methods. The enthalpy change may also be calculated from empirical data using thermochemical cycles, bond dissociation energies, $D$(R–H), and electron affinities, EA(R).[33] Furthermore, the entropy change can be calculated with the partition function thus leading to the Gibbs free energy change, $\Delta G_a$. The entropy changes for the different compounds are near 100 J mol$^{-1}$ and do not change greatly.

**Table 3** Correlations for individual sets (outliers dropped) $\Delta G_a = aV_{mc} + b\pi_I + c\varepsilon_B + dq_- + e\varepsilon_A + fq_+ + g$

| Coeff.<br>±<br>*t*-stat.<br>*P*(2-tail)<br>VIF | *a* | *b* | *c* | *d* | *e* | *f* | *g* | *n* | *R* | SD | *F* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Acids | −17.6 | n/s | n/s | 676 | −1654 | −1710 | 1751 | 27 | 0.937 | 8.4 | 39 |
| | 8.1 | | | 89 | 359 | 233 | 81 | | | | |
| | 2.16 | | | 7.53 | 4.60 | 7.34 | 21.6 | | | | |
| | 0.042 | | | 0.000 | 0.000 | 0.000 | 0.000 | | | | |
| | 1.29 | | | 1.25 | 1.20 | 1.31 | | | | | |
| Alcohols | −33.2 | n/s | n/s | 900 | n/s | −1708 | 1582 | 20 | 0.994 | 3.6 | 448 |
| | 2.4 | | | 45 | | 160 | 40 | | | | |
| | 14.1 | | | 20.0 | | 10.7 | 39.3 | | | | |
| | 0.000 | | | 0.000 | | 0.000 | 0.000 | | | | |
| | 1.08 | | | 1.66 | | 1.75 | | | | | |
| Silanols | n/s | n/s | n/s | 388 | n/s | −2271 | 1678 | 17 | 0.983 | 7.7 | 206 |
| | | | | 9.7 | | 115 | 58 | | | | |
| | | | | 4.00 | | 19.8 | 44.6 | | | | |
| | | | | 0.001 | | 0.000 | 0.000 | | | | |
| | | | | 1.20 | | 1.20 | | | | | |
| Anilines | n/s | n/s | n/s | n/s | −998 | n/s | 536 | 13 | 0.850 | 2.1 | 29 |
| | | | | | 187 | | 34 | | | | |
| | | | | | 5.35 | | 16.0 | | | | |
| | | | | | 0.000 | | 0.000 | | | | |
| | | | | | 1.00 | | | | | | |
| Oximes | n/s | n/s | n/s | n/s | −6011 | n/s | 2541 | 7 | 0.831 | 39 | 11 |
| | | | | | 1801 | | 328 | | | | |
| | | | | | 3.34 | | 7.76 | | | | |
| | | | | | 0.021 | | 0.001 | | | | |
| | | | | | 1.00 | | | | | | |
| Hydrocarbons | n/s | n/s | −1197 | n/s | −871.6 | n/s | 342.3 | 15 | 0.982 | 6.5 | 159 |
| | | | 226 | | 167 | | 17.1 | | | | |
| | | | 5.29 | | 4.23 | | 20.0 | | | | |
| | | | 0.000 | | 0.001 | | 0.000 | | | | |
| | | | 3.78 | | 3.78 | | | | | | |
| Small hydrocarbons | n/s | n/s | n/s | n/s | −4950 | n/s | 2467 | 7 | 0.979 | 8.0 | 116 |
| | | | | | 460 | | 76 | | | | |
| | | | | | 10.8 | | 32.4 | | | | |
| | | | | | 0.000 | | 0.000 | | | | |
| | | | | | 1.00 | | | | | | |

The $\Delta H_a$ and $\Delta G_a$ values are capable of being determined with four figure precision. Many data analysed by QSAR methods are reliable to two or three digits only.

Molecular computations also can produce $\Delta H_a$. Examples include *ab initio* calculations: (also MNDO) for nine small molecules of the type, $XH_n$, by Gordon and co-workers;[34] at G2 level for 23 small molecules by Smith and Radom;[35] for four α-ketoaldoximes by Bouchoux and colleagues;[36] and 12 small allyls, ethenes and cyclopropanes by Dahlke and Kass.[37] There is very good agreement with experiment and theory.

## Procedure

This study applies the TLSER descriptors to the gas phase acidity for some sets of carboxylic acids,[38] alcohols and anilines,[39] silanols,[40] hydrocarbons,[33] oximes[36] and small, unsaturated and cyclic hydrocarbons.[37] Experimental data was selected from the compilation of Lias and co-workers[41] and the previous references.

Molecular geometries were optimized and TLSER descriptors were calculated using the MNDO algorithm contained in MOPAC.[42,43] The molecular volume for the optimized geometry was determined using the algorithm of Hopfinger.[44] The in house developed molecular modelling package, MMADS, was used to construct and view all molecular structures.[45] Multilinear regression analysis [using MYSTAT (Systat, Evanston, Il, USA)] was used to obtain the coefficients in the correlation equation.

The correlation equations were selected based on the coefficients being significant at the 0.95 level ('large' *t*-statistic) or higher, the correlation coefficient, *R*, as large as possible, small cross correlation [small variance inflation factor (VIF)] and a minimum number of outliers. The VIF is defined as $1/(1 - R^2)$ where *R* is the correlation coefficient of one variable against the others; small (closer to one) values imply small cross correlation.[46] Outliers were taken as compounds whose calculated values were three or more standard deviations from the mean.

## Results

Tables 2, 3 and 4 summarize the correlation equations. Each coefficient is accompanied by its standard error (±), *t*-statistic (*t*-stat.), 2-tail probability *P*(2-tail) and the variance inflation factor (VIF) to indicate the quality of the 'fit'. There is some redundancy since small standard error accompanies a large *t*-statistic and small probability.

Tables 2 and 3 are for the individual classes of compounds:

**Table 4** Correlations for some continued sets (outliers dropped)[a] $\Delta G_a = aV_{mc} + b\pi_1 + c\varepsilon_B + dq_- + e\varepsilon_A + fq_+ + g$

| Coeff.<br>±<br>*t*-stat.<br>*P*(2-tail)<br>VIF | *a* | *b* | *c* | *d* | *e* | *f* | *g* | *n* | *R* | SD | *F* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Acids + alcohols | −24.14 | n/s | n/s | 366.7 | −528.7 | −3548 | 2155 | 46 | 0.994 | 7.7 | 806 |
| | 4.31 | | | 52.8 | 196.4 | 174 | 18 | | | | |
| | 5.60 | | | 6.94 | 2.69 | 20.4 | 119 | | | | |
| | 0.000 | | | 0.000 | 0.010 | 0.000 | 0.000 | | | | |
| | 1.22 | | | 1.79 | 5.65 | 7.87 | | | | | |
| Acids + alcohols | −27.28 | n/s | n/s | 429.1 | n/s | −3967 | 2135 | 46 | 0.993 | 8.3 | 933 |
| | 4.45 | | | 50.9 | | 84 | 18 | | | | |
| | 6.13 | | | 8.43 | | 47.2 | 120 | | | | |
| | 0.000 | | | 0.000 | | 0.000 | 0.000 | | | | |
| | 1.13 | | | 1.44 | | 1.60 | | | | | |
| Acids + alcohols + silanols | n/s | n/s | n/s | 67.32 | −1502 | −2407 | 2162 | 60 | 0.982 | 11 | 515 |
| | | | | 11.09 | 186 | 142 | 19 | | | | |
| | | | | 6.07 | 8.09 | 17.0 | 113 | | | | |
| | | | | 0.000 | 0.000 | 0.000 | 0.000 | | | | |
| | | | | 1.15 | 2.71 | 2.88 | | | | | |
| Acids + alcohols + anilines | n/s | n/s | 1087 | n/s | −3551 | −856.0 | 2059 | 59 | 0.983 | 11 | 541 |
| | | | 264 | | 127 | 61.7 | 39 | | | | |
| | | | 4.11 | | 28.0 | 13.9 | 39.2 | | | | |
| | | | 0.000 | | 0.000 | 0.000 | 0.000 | | | | |
| | | | 3.54 | | 1.47 | 2.88 | | | | | |
| Alcohols + silanols + anilines | n/s | 841.3 | n/s | n/s | −3019 | −631.3 | 2004 | 46 | 0.976 | 9.9 | 282 |
| | | 141.7 | | | 118 | 55.0 | 24 | | | | |
| | | 5.94 | | | 25.6 | 11.5 | 83.0 | | | | |
| | | 0.000 | | | 0.000 | 0.000 | 0.000 | | | | |
| | | 3.01 | | | 1.54 | 2.38 | | | | | |
| Acids + alcohols + silanols + anilines | n/s | n/s | 1160 | n/s | −3439 | −835.6 | 2027 | 72 | 0.975 | 13 | 432 |
| | | | 226 | | 128 | 57.4 | 36 | | | | |
| | | | 5.14 | | 26.9 | 14.6 | 55.9 | | | | |
| | | | 0.000 | | 0.000 | 0.000 | 0.000 | | | | |
| | | | 2.59 | | 1.30 | 2.09 | | | | | |
| Acids + alcohols + silanols + anilines | n/s | 903.8 | n/s | 67.34 | −3372 | −950.8 | 2081 | 74 | 0.970 | 14 | 271 |
| | | 195.8 | | 15.59 | 155 | 59.3 | 26 | | | | |
| | | 4.62 | | 4.32 | 21.7 | 16.0 | 79.9 | | | | |
| | | 0.000 | | 0.000 | 0.000 | 0.000 | 0.000 | | | | |
| | | 3.08 | | 1.83 | 1.62 | 1.91 | | | | | |

[a] For a list of outliers, see Table 7.

acids, alcohols, silanols, anilines, oximes, hydrocarbons and some small hydrocarbons made up of propenes, ethenes and cyclopropanes. Table 2 includes all compounds while Table 3 has the outliers removed. Both tables were included for purposes of comparison. Table 4 contains equations for combined classes of acids, alchols, silanols and anilines; the outliers were removed. The first two rows (both acids and alcohols combined) are included for comparison with the individual acid and alcohol sets. The last two rows, each for the combined acids, alcohols, anilines, and silanols were included for comparison. Table 5 contains the TLSER descriptors for compounds used in Table 4 while Table 6 contains the experimental, calculated and residual acidity measure values for the 72 compound case of these acids, alcohols, anilines and silanols. Table 7 lists the outliers for the compounds sets in Table 4. Table 8 contains the hydrocarbons, oximes and small hydrocarbons along with their descriptors. There are 108 compounds in all.

Since most of the experimental parameters are measured to four significant figures the TLSER descriptors and coefficients are similarly reported.

The combined sets were those for which $\Delta G_a$ was available. The enthalpy change, $\Delta H_a$, often used as the measure of gas phase acidity, was not readily available for the acids and silanols. The hydrocarbon set (ref. 33) used the activation energy for the MgO-catalysed protium and deuterium exchange between XH(g) and $D_2$(g). This was included in this paper because there is a strong linear correlation of activation enthalpy, $\Delta H^{\pm} = E_{ac} - 2RT$, with $\Delta H_a$ for a subset of compounds for which the experimental values were known.

## Discussion

Examination of the tables shows the general trends and good correlations ($0.823 < R < 0.994$) generated by the TLSER descriptors; the combined, hence larger, sets have $R$ of the order of 0.970 or higher. Not all terms are significant for the various compound sets. The retained terms are significant at the 0.95 level or higher; most are significant beyond the 0.99 level. Standard error of the estimated (SD) values are near the estimated experimental uncertainties, typically ±2–8 kJ mol$^{-1}$. Some correlations for individual classes have SD values around a third to a half of the upper value. This could mean that the equations are artifacts. However, that would seem to be precluded since none are lower than ±2 kJ mol$^{-1}$ and the coefficients have strong statistical significance.

**Table 5** Acids, alcohols, anilines and silanols with descriptors

| Compound | $V_{mc}$ | $\pi_I$ | $\varepsilon_B$ | $q_-$ | $\varepsilon_A$ | $q_+$ | $\Delta G_a$ |
|---|---|---|---|---|---|---|---|
| Carboxylic acids[a] | | | | | | | |
| ethanoic | 0.5949 | 0.0843 | 0.1298 | 0.3650 | 0.1696 | 0.2161 | 1429 |
| propanoic | 0.7067 | 0.0970 | 0.1304 | 0.3665 | 0.1693 | 0.2161 | 1424 |
| butanoic | 0.8923 | 0.0986 | 0.1305 | 0.3648 | 0.1691 | 0.2151 | 1420 |
| cyclopropylmethanoic | 0.7917 | 0.1043 | 0.1313 | 0.3655 | 0.1690 | 0.2151 | 1420 |
| cyclopropylethanoic | 0.9772 | 0.1039 | 0.1332 | 0.3612 | 0.1689 | 0.2151 | 1418 |
| (2-$CH_3$-cyclopropyl)methanoic | 0.9775 | 0.1047 | 0.1321 | 0.3688 | 0.1699 | 0.2176 | 1418 |
| 3-$CH_3$-but-2-enoic | 0.9945 | 0.1079 | 0.1419 | 0.3664 | 0.1787 | 0.2147 | 1415 |
| (*E*)-pent-2-enoic | 0.9935 | 0.1081 | 0.1402 | 0.3787 | 0.1796 | 0.2157 | 1415 |
| methanoic | 0.3561 | 0.0874 | 0.1282 | 0.3695 | 0.1685 | 0.2161 | 1415 |
| 3,3-$(CH_3)_2$-butanoic | 1.2330 | 0.1011 | 0.1311 | 0.3648 | 0.1692 | 0.2160 | 1415 |
| (*E*)/(*Z*) (mix)-but-2-enoic[e] | 0.8102 | 0.1088 | 0.1403 | 0.3775 | 0.1797 | 0.2156 | 1411 |
| pent-4-enoic | 0.9901 | 0.1054 | 0.1434 | 0.3661 | 0.1692 | 0.2155 | 1411 |
| prop-2-enoic | 0.6421 | 0.1029 | 0.1376 | 0.3794 | 0.1787 | 0.2154 | 1411 |
| (1-$CH_3$-cyclopropyl)ethanoic | 0.9775 | 0.1047 | 0.1321 | 0.3688 | 0.1699 | 0.2176 | 1410 |
| 2-$CH_3$-prop-2-enoic | 1.7968 | 0.1078 | 0.1402 | 0.3663 | 0.1769 | 0.2160 | 1407 |
| (bicyclo[1.1.1]pentyl)methanoic | 1.0400 | 0.1098 | 0.1326 | 0.3662 | 0.1694 | 0.2145 | 1405 |
| $CH_3O$-ethanoic | 0.7841 | 0.0991 | 0.1308 | 0.3307 | 0.1697 | 0.2176 | 1402 |
| 3,3-$(CH_3)_2$-pent-2-enoic | 1.3478 | 0.1040 | 0.1440 | 0.3662 | 0.1692 | 0.2157 | 1402 |
| phenylethanoic | 1.2624 | 0.1208 | 0.1483 | 0.3011 | 0.1830 | 0.1815 | 1398 |
| 4-oxopentanoic | 1.0617 | 0.1011 | 0.1360 | 0.3690 | 0.1736 | 0.2171 | 1396 |
| fluoroethanoic | 0.5631 | 0.0908 | 0.1278 | 0.3293 | 0.1733 | 0.2231 | 1385 |
| 4-$F_3$-butanoic | 0.9895 | 0.0890 | 0.1253 | 0.3599 | 0.1742 | 0.2226 | 1379 |
| chloroethanoic | 0.6801 | 0.1014 | 0.1267 | 0.3262 | 0.1751 | 0.2231 | 1376 |
| bromoethanoic | 0.7372 | 0.1094 | 0.1282 | 0.3314 | 0.1783 | 0.2217 | 1370 |
| 2-oxopropanoic | 0.7030 | 0.1003 | 0.1340 | 0.3424 | 0.1786 | 0.2213 | 1366 |
| difluoroethanoic | 0.5936 | 0.0871 | 0.1242 | 0.3265 | 0.1773 | 0.2277 | 1354 |
| dichloroethanoic | 0.8278 | 0.1090 | 0.1242 | 0.3300 | 0.1846 | 0.2285 | 1347 |
| Alcohols and phenols[b] | | | | | | | |
| methanol | 0.3647 | 0.0860 | 0.1314 | 0.3291 | 0.1402 | 0.1803 | 1559 |
| ethanol[d] | 0.5423 | 0.0927 | 0.1326 | 0.3235 | 0.1429 | 0.1800 | 1551 |
| propanol[d] | 0.7237 | 0.0955 | 0.1326 | 0.3246 | 0.1442 | 0.1803 | 1543 |
| butanol[a] | 0.9082 | 0.0969 | 0.1326 | 0.3249 | 0.1449 | 0.1804 | 1541 |
| 2-propanol | 0.7148 | 0.0962 | 0.1335 | 0.3197 | 0.1448 | 0.1780 | 1538 |
| 3-$CH_3$-butanol | 1.0752 | 0.0991 | 0.1326 | 0.3254 | 0.1461 | 0.1802 | 1535 |
| 2-$CH_3$-propanol | 0.9106 | 0.0964 | 0.1327 | 0.3245 | 0.1457 | 0.1807 | 1535 |
| 2-$CH_3$-propan-2-ol | 0.8929 | 0.0976 | 0.1342 | 0.3182 | 0.1442 | 0.1765 | 1534 |
| 2,2-$(CH_3)_2$-butan-4-ol[a] | 1.2477 | 0.0998 | 0.1337 | 0.3234 | 0.1463 | 0.1783 | 1531 |
| $CH_3O$-ethanol | 0.8028 | 0.0966 | 0.1354 | 0.3545 | 0.1462 | 0.1843 | 1530 |
| 2,2-$(CH_3)_2$-propanol | 1.0673 | 0.0995 | 0.1327 | 0.3257 | 0.1458 | 0.1820 | 1528 |
| 2,2-$(CH_3)_2$-butan-3-ol[a] | 1.2565 | 0.0993 | 0.1338 | 0.3240 | 0.1478 | 0.1798 | 1525 |
| 3,3-$(CH_3)_2$-butan-2-ol | 1.2512 | 0.0997 | 0.1338 | 0.3245 | 0.1481 | 0.1803 | 1523 |
| phenylmethanol[d] | 1.0933 | 0.1190 | 0.1527 | 0.3252 | 0.1758 | 0.1825 | 1520 |
| 2,2-$(CH_3)_2$-pentanol | 1.4055 | 0.1022 | 0.1355 | 0.3271 | 0.1486 | 0.1807 | 1519 |
| 2,2,4-$(CH_3)_3$-pentanol | 1.6299 | 0.0995 | 0.1359 | 0.3237 | 0.1496 | 0.1795 | 1514 |
| 2,2,4,4-$(CH_3)_4$-pentanol | 1.7828 | 0.1013 | 0.1364 | 0.3298 | 0.1496 | 0.1813 | 1509 |
| 2,2-$F_2$-ethanol | 0.6115 | 0.0835 | 0.1259 | 0.3057 | 0.1545 | 0.1924 | 1505 |
| 2,2,2-$F_3$-ethanol | 0.6464 | 0.0782 | 0.1212 | 0.3096 | 0.1644 | 0.2035 | 1493 |
| 4-$CH_3$-phenol | 1.0874 | 0.1230 | 0.1572 | 0.2496 | 0.1764 | 0.1936 | 1446 |
| phenol | 0.9156 | 0.1222 | 0.1567 | 0.2481 | 0.1756 | 0.1928 | 1441 |

One (or both) of the acidity descriptors, $\varepsilon_A$ and $q_+$, is (are) included in each correlation (with one exception in Table 2, discussed later) as would be expected. Furthermore their signs are negative, indicating that increasing acidity decreases $\Delta H_a$, as it should. The sign on the $q_-$ terms are positive indicating increasing basicity increases $\Delta H_a$, again, as expected.

The $\varepsilon_B$ descriptor occurs in the hydrocarbon case, the acid, alcohol and aniline case and one of the combined ($N = 72$) sets of acids, alcohols, silanols and anilines. Its sign is positive, in the combined case sets. Again this is physically consistent. However, for the hydrocarbons the sign is negative. Ref. 33 suggests an explanation since this actually involves $E_{ac}$, the activation energy, for a deuterium exchange process over a solid catalyst, MgO. The bifunctionality of the MgO permits transition states involving O–D, O–H and Mg–R. Increasing basicity, $\varepsilon_B$, would favour lowering of the activation energy. Another aspect of the hydrocarbons is that since they lack a 'real' basic site they must use the molecular orbital basicity. The hydrocarbon cases in Tables 2 and 3 have $\varepsilon_B$ as the only significant parameter. This could mean the HBAB (covalent electron donating functionality) of the hydrocarbon can interact with the acidity function of the MgO as discussed in the previous paragraph and account for most of the variance.

The volume terms are significant, with negative signs, for acids and alcohols. This might be associated with an inductive effect which is expected to be larger for larger molecules.

For the combined acid and alcohol set a statistically acceptable correlation equation was obtained involving the polarizability index (along with $V_{mc}$, $\varepsilon_A$ and $q_+$) with a negative sign; again, this is physically reasonable. The detailed equation is not reported since it had lower $R$ and higher SD values. Of the two acid and alcohol correlations in Table 4 the second is better than the first because of the smaller VIF values indicating smaller cross correlation of the coefficients. When alcohols, silanols and anilines are combined $\pi_I$ is found to be significant (along with $\varepsilon_A$ and $q_+$) but with a positive sign; but, it was not significant for alcohols and silanols combined. Inclusion of the anilines, quite polarizable, introduces a wider

Table 5 (*continued*)

| Compound | $V_{mc}$ | $\pi_I$ | $\varepsilon_B$ | $q_-$ | $\varepsilon_A$ | $q_+$ | $\Delta G_a$ |
|---|---|---|---|---|---|---|---|
| Anilines[b] | | | | | | | |
| 4-$CH_3$ | 1.1433 | 0.1226 | 0.1542 | 0.2220 | 0.1775 | 0.1012 | 1510 |
| 4-$CH_3O$ | 1.2219 | 0.1240 | 0.1569 | 0.2848 | p.1769 | 0.0956 | 1509 |
| 3-$CH_3$ | 1.1395 | 0.1244 | 0.1581 | 0.2299 | 0.1758 | 0.1146 | 1508 |
| aniline | 0.9693 | 0.1245 | 0.1580 | 0.2279 | 0.1748 | 0.1133 | 1505 |
| 4-F | 1.0026 | 0.1232 | 0.1568 | 0.2270 | 0.1793 | 0.1171 | 1499 |
| 2-F | 1.0084 | 0.1217 | 0.1551 | 0.2319 | 0.1791 | 0.1208 | 1495 |
| 3-$CH_3S$ | 1.3307 | 0.1177 | 0.1503 | 0.2226 | 0.1806 | 0.0998 | 1492 |
| 3-F | 1.0041 | 0.1222 | 0.1551 | 0.2322 | 0.1791 | 0.1210 | 1489 |
| 2,4-$F_2$ | 1.0294 | 0.1223 | 0.1546 | 0.2263 | 0.1832 | 0.1318 | 1486 |
| 4-Cl | 1.1304 | 0.1271 | 0.1557 | 0.2309 | 0.1795 | 0.1202 | 1486 |
| 3-Cl | 1.1243 | 0.1272 | 0.1549 | 0.2289 | 0.1794 | 0.1192 | 1480 |
| 3-$CF_3$ | 1.2365 | 0.1116 | 0.1453 | 0.2339 | 0.1857 | 0.1025 | 1472 |
| 4-$CF_3$ | 1.2374 | 0.1150 | 0.1507 | 0.2367 | 0.1841 | 0.1291 | 1457 |
| Silanols[c] | | | | | | | |
| $(CH_3)H_2SiOH$ | 0.6459 | 0.0821 | 0.1348 | 0.6562 | 0.1531 | 0.2026 | 1477 |
| $(CH_3)_2HSiOH$ | 0.8213 | 0.0864 | 0.1362 | 0.6371 | p.1547 | 0.1988 | 1477 |
| $(CH_3)_3SiOH$ | 1.0180 | 0.0959 | 0.1542 | 0.7351 | 0.1758 | 0.2315 | 1477 |
| $(CH_3)_2(C_2H_5)SiOH$ | 1.1531 | 0.0937 | 0.1366 | 0.6253 | 0.1758 | 0.1960 | 1477 |
| $(CH_3)(C_2H_5)_2SiOH$ | 1.3324 | 0.0953 | 0.1379 | 0.6261 | 0.1560 | 0.1971 | 1477 |
| $(C_2H_5)_3SiOH$ | 0.5351 | 0.0951 | 0.1385 | 0.6265 | 0.1560 | 0.1977 | 1477 |
| $(C_6H_5)(CH_3)_2SiOH$ | 1.5259 | 0.1109 | 0.1522 | 0.6286 | 0.1759 | 0.1960 | 1477 |
| $H_3SiOH$ | 0.4808 | 0.0719 | 0.1346 | 0.6699 | 0.1518 | 0.2046 | 1473 |
| $(C_6H_5)(CH_3)HSiOH$ | 1.3663 | 0.1106 | 0.1522 | 0.6431 | 0.1761 | 0.1992 | 1464 |
| $(C_6H_5)H_2SiOH$ | 1.2054 | 0.1098 | 0.1525 | 0.6648 | 0.1758 | 0.2048 | 1456 |
| $F(CH_3)_2SiOH$ | 0.8486 | 0.0803 | 0.1334 | 0.6550 | 0.1557 | 0.2105 | 1456 |
| $(CH_3O)_2(CH_3)SiOH$ | 1.1702 | 0.0876 | 0.1523 | 0.6839 | 0.1577 | 0.2157 | 1456 |
| $(CH_3O)_3SiOH$ | 1.2201 | 0.0887 | 0.1333 | 0.6807 | 0.1629 | 0.2210 | 1456 |
| $(CH_3O)_2HSiOH$ | 0.9962 | 0.0853 | 0.1338 | 0.6794 | 0.1606 | 0.2164 | 1456 |
| $F_2(CH_3)SiOH$ | 0.7015 | 0.0667 | 0.1187 | 0.6717 | 0.1623 | 0.2229 | 1423 |
| $F_3SiOH$ | 0.5653 | 0.0441 | 0.1122 | 0.6776 | 0.1636 | 0.2412 | 1393 |
| $Cl_3SiOH$ | 0.9309 | 0.1010 | 0.1128 | 0.6469 | 0.1968 | 0.2645 | 1326 |

[a] Ref. 34. [b] Ref. 35. [c] Ref. 36. [d] Ref. 37. [e] Parameters for *E*, see discussion.

range of $\pi_I$; however, there is no ready explanation for the sign. More importantly, the combined set of acids, alcohols, anilines and silanols gave a correlation with only four outliers out of 78 compounds when $\pi_I$ (along with $q_-$, $\varepsilon_A$, and $q_+$) was included; again, it had a positive coefficient. The other combined case reported in Table 4 for the four classes of compounds had six outliers.

Table 5 includes data for what is labelled as (*E*)/(*Z*)-but-2-enoic acid. That was the mixture for which the value was obtained. The descriptors listed here are for the *E* form; there is only a small difference in the parameters for the two isomers. Either set of parameters gave essentially the same correlation for the combined cases of acids, alcohols, anilines and silanols with the *R* value being identical to three digits.

Tables 2 and 3 show that the individual classes of compounds fall into two categories. Those with $q_+$ significant, the acids, alcohols and silanols, and those with $q_+$ not significant, the anilines, oximes, hydrocarbons and small hydrocarbons. (The one exception is the hydrocarbons in Tables 2 and 3 for which $\varepsilon_B$ is the only significant parameter; it was discussed earlier.) The first type have OH groups and, thus, a readily identifiable acidic proton. The second type, except for the oximes, do not have readily identifiable acidic protons. A possible explanation is the small oxime sample size, seven, with four of these being α-ketoaldoximes.

Comparison of the individual correlations (Tables 2 and 3) shows other physically reasonable results. The expected acidity order, alcohols < acids < silanols, is consistent with the order, alcohols > acids > silanols, of the coefficient, *d* of the $q_-$ descriptor (basicity). Decreasing *d* means increasing acidity. On the molecular level Table 5 shows that the oxygens are more negative ($q_-$) in the silanols than in the acids; charge separation can be accommodated better by Si than C. Corresponding observations result from the order of the coefficient, *f*, of the $q_+$ descriptor (acidity). Smaller *f* values (algebraically) indicate greater acidity.

*Outliers.*—Methoxyethanol was the only outlier in the alcohols; however, if the OH oxygen charge is used instead of that of the more negative $CH_3O$ oxygen, methoxyethanol is no longer an outlier and *R* increases to 0.992. Trimethylsilanol is the only outlier for that class; its $q_+$ value needs to be larger and $q_-$ smaller to provide a better fit. Benzene is the outlier in the hydrocarbon set; its calculated value is smaller than the experimental one. If its $\varepsilon_B$ [*E*(HOMO) magnitude] value were smaller there would be better agreement with experiment. In fact, based on $\varepsilon_B$ its calculated value is close to that of thiophene while its experimental value is more than twice as large. Ethene was the outlier for the small hydrocarbons composed of propenes, ethenes and cyclopropanes. This was a small set, only eight; ethene is the smallest member. Larger sets of related hydrocarbons might give different results. The results for the small sets, oximes (seven) and small hydrocarbons (eight), are included to give some indication as to their correlation; hence they are far from being definitive.

There is one common outlier, $CF_3CH_2OH$, in the combined sets. When anilines are not included, the calculated value is too small and $q_+$ contributes most to the equation. If it were smaller, the calculated value would be closer to that for experiment for this alcohol. The three fluorines on the alcohol could explain the larger $q_+$ value; it is larger than for ethanol. When anilines are combined, the calculated value is too large and the ε's contribute most. Again the presence of the fluorine seems to account for this; relative to ethanol, it increases the $\varepsilon_B$ and lowers $\varepsilon_A$, thus making the calculated value too large. MNDO is noted for its inability to handle polyfluorinated

**Table 6** $\Delta G_\alpha$ Values for combined set of 72 acids, alcohols, anilines, and silanols in Table 4

| Compound | Expmt. | Calc. | Residual | Compound | Expmt. | Calc. | Residual |
|---|---|---|---|---|---|---|---|
| Carboxylic acids | | | | | | | |
| ethanoic | 1429 | 1413 | 15 | 2,2-$(CH_3)_2$-propanol | 1528 | 1527 | 1 |
| propanoic | 1424 | 1415 | 9 | 2,2-$(CH_3)_2$-butan-3-ol | 1525 | 1523 | 2 |
| butanoic | 1420 | 1417 | 3 | 3,3-$(CH_3)_2$-butan-2-ol | 1523 | 1522 | 1 |
| (cyclopropyl)methanoic | 1420 | 1419 | 2 | 2,2-$(CH_3)_2$-pentanol | 1519 | 1522 | −3 |
| (cyclopropyl)ethanoic | 1418 | 1421 | −3 | 2,2,4-$(CH_3)_3$-pentanol | 1514 | 1520 | −6 |
| (2-$CH_3$-cyclopropyl)methanoic | 1418 | 1414 | 4 | 2,2,4,4-$(CH_3)_4$-pentanol | 1509 | 1519 | −1 |
| 3-$CH_3$-but-2-enoic | 1415 | 1397 | 18 | 2,2-$F_2$-ethanol | 1505 | 1481 | 25 |
| (*E*)-pent-2-enoic | 1415 | 1391 | 24 | 4-$CH_3$-phenol | 1446 | 1441 | 6 |
| methanoic | 1415 | 1415 | −0 | phenol | 1441 | 1444 | −2 |
| 3,3-$(CH_3)_2$-butanoic | 1415 | 1416 | −1 | | | | |
| (*E*)/(*Z*)-but-2-enoic | 1411 | 1391 | 20 | Anilines | | | |
| pent-4-enoic | 1411 | 1431 | −20 | 4-$CH_3$ | 1510 | 1511 | −0 |
| prop-2-enoic | 1411 | 1392 | 19 | 4-$CH_3O$ | 1509 | 1520 | −11 |
| (1-$CH_3$-cyclopropyl)methanoic | 1410 | 1414 | −3 | 3-$CH_3$ | 1507 | 1510 | −2 |
| 2-$CH_3$-prop-2-enoic | 1407 | 1401 | 6 | aniline | 1505 | 1514 | −9 |
| (bicyclo[1.1.1]pentyl)methanoic | 1405 | 1419 | −14 | 4-F | 1499 | 1494 | 5 |
| $CH_3O$-ethanoic | 1402 | 1413 | −11 | 2-F | 1495 | 1490 | 5 |
| 3,3-$(CH_3)_2$-pent-2-enoic | 1402 | 1432 | −30 | 3-$CH_3S$ | 1492 | 1497 | −4 |
| phenylethanoic | 1398 | 1418 | −20 | 3-F | 1489 | 1490 | −1 |
| 4-oxopentanoic | 1396 | 1406 | −10 | 2,4-$F_2$ | 1486 | 1466 | 20 |
| fluoroethanoic | 1385 | 1393 | −8 | 4-Cl | 1486 | 1490 | −3 |
| 4,4,4-$F_3$-butanoic | 1379 | 1387 | −8 | 3-Cl | 1480 | 1490 | −10 |
| chloroethanoic | 1376 | 1385 | −9 | 3-$CF_3$ | 1472 | 1471 | 1 |
| bromoethanoic | 1370 | 1377 | −7 | 4-$CF_3$ | 1457 | 1461 | −3 |
| 2-oxopropanoic | 1366 | 1383 | −17 | | | | |
| difluoroethanoic | 1354 | 1370 | −16 | Silanols | | | |
| dichloroethanoic | 1347 | 1345 | 2 | $CH_3H_2SiOH$ | 1477 | 1487 | −10 |
| | | | | $(CH_3)_2HSiOH$ | 1477 | 1487 | −10 |
| Alcohols and phenols | | | | $(CH_3)(C_2H_5)_2SiOH$ | 1477 | 1486 | −9 |
| methanol | 1559 | 1546 | 13 | $(C_2H_5)_3SiOH$ | 1477 | 1484 | −7 |
| ethanol | 1551 | 1539 | 12 | $H_3SiOH$ | 1473 | 1490 | −17 |
| propanol | 1543 | 1534 | 9 | $(C_6H_5)(CH_3)HSiOH$ | 1464 | 1431 | 33 |
| butanol | 1541 | 1532 | 10 | $(C_6H_5)H_2SiOH$ | 1456 | 1428 | 28 |
| 2-propanol | 1538 | 1535 | 3 | $F(CH_3)_2SiOH$ | 1456 | 1470 | −14 |
| 2-$CH_3$-propanol | 1535 | 1529 | 6 | $(CH_3O)_2(CH_3)SiOH$ | 1456 | 1481 | −25 |
| 3-$CH_3$-butan-1-ol | 1535 | 1528 | 8 | $(CH_3O)_3SiOH$ | 1456 | 1436 | 20 |
| 2-$CH_3$-propan-2-ol | 1534 | 1539 | −5 | $(CH_3O)_2HSiOH$ | 1456 | 1449 | 7 |
| 2,2-$(CH_3)_2$-butan-4-ol | 1531 | 1530 | 1 | $F_2(CH_3)SiOH$ | 1423 | 1420 | 3 |
| $CH_3O$-ethanol | 1530 | 1527 | 3 | $F_3SiOH$ | 1393 | 1393 | 1 |

**Table 7** Outliers for combined sets in Table 4

| System | Compounds |
|---|---|
| Acids + alcohols | $C_6H_5CH_2CO_2H$, $CF_3CH_2OH$ |
| Acids + alcohols + silanols | $C_6H_5CH_2CO_2H$, $CF_3CH_2OH$, $C_6H_5CH_2OH$, $(CH_3)_3SiOH$, $Cl_3SiOH$ |
| Acids + alcohols + anilines | $CF_3CH_2OH$, $C_6H_5CH_2OH$ |
| Alcohols + silanols + anilines | $CF_3CH_2OH$, $C_6H_5CH_2OH$, $(CH_3)_3SiOH$, $(CH_3)_2(C_2H_5)SiOH$, $(CH_3)_2(C_6H_5)SiOH$ |
| Acids + alcohols + silanols + anilines | $CF_3CH_2OH$, $C_6H_5CH_2OH$, $(CH_3)_3SiOH$, $Cl_3SiOH$, $(CH_3)_2(C_2H_5)SiOH$, $(CH_3)_2(C_6H_5)SiOH$ |
| Acids + alcohols + silanols + anilines | $CF_3CH_2OH$, $C_6H_5CH_2OH$, $(CH_3)_3SiOH$, $(CH_3)_2(C_2H_5)SiOH$ |

species. It is of interest to note that difluoroethanol is not an outlier.

The acid, $C_6H_5CH_2CO_2H$, is an outlier when the acids, alcohols and silanols are combined; its calculated value is too high. A larger $q_+$, the most significant term, would bring the calculated value into line. In this connection, $C_6H_5CH_2OH$, is an outlier for combinations of three or more classes; calculated values are too high. Here at least two parameters make strong contributions so the interpretation is complicated. However, a larger $q_+$ would bring the calculated value in line. To a rough approximation the phenyl group seems to coincide with a smaller $q_+$ than expected compared to the other members of the set.

Trimethylsilanol is an outlier in all correlations; calculated values are too large. It has the largest $q_-$ value in the silanols, 0.735; this would contribute to a higher $\Delta G_\alpha$ in three of the sets but not the others. (It should be noted that the formal charges on the Si are greater than one for the 18 silanols.) In the largest set, silanols with two methyls along with a methyl, ethyl or phenyl are outliers. Again, calculated values exceed the experimental. Their formal charges are less than that for trimethylsilanol; consequently, a general explanation for these outliers is not apparent. Trichlorosilanol is also an outlier in several cases. It has the largest $q_+$ value; that is not the problem since it helps make $\Delta G_\alpha$ smaller.

In general, the presence of an outlier can indicate that a different mechanism is involved. This could apply when a kinetic process is involved as is often the case in QSAR. However, that would not apply for a thermodynamic quantity. An error in the data could also explain an outlier. A possible explanation for an outlier when using computational descriptors could be inadequacies in the model used for molecular computations.

## Conclusions

The TLSER descriptors provide good to very good correlation equations for the gas phase acidities for representative sets, including combinations, of acids, alcohols, silanols, anilines,

**Table 8** Hydrocarbons, oximes and small hydrocarbons with descriptors

| Compound | $V_{mc}$ | $\pi_1$ | $\varepsilon_B$ | $q_-$ | $\varepsilon_A$ | $q_+$ | $\Delta H_a$[a] |
|---|---|---|---|---|---|---|---|
| Hydrocarbons[b] | | | | | | | |
| propane | 0.6527 | 0.0957 | 0.1222 | 0.0346 | 0.1430 | 0.0048 | 78.12 |
| 2,2-$(CH_3)_2$-propane | 1.0015 | 0.0991 | 0.1244 | 0.1258 | 0.1429 | −0.0069 | 73.14 |
| 2-$(CH_3)$-butane | 1.0120 | 0.0986 | 0.1254 | 0.0662 | 0.145 | 0.0156 | 72.55 |
| ethane | 0.4776 | 0.0909 | 0.1189 | 0.0071 | 0.1408 | −0.0046 | 69.29 |
| cyclohexane | 1.0650 | 0.1054 | 0.1284 | 0.0111 | 0.1460 | 0.0059 | 64.89 |
| cyclopentane | 0.8926 | 0.1024 | 0.1249 | 0.0169 | 0.1461 | 0.0090 | 62.80 |
| cyclobutane | 0.7285 | 0.1005 | 0.1276 | 0.0340 | 0.1452 | 0.0170 | 60.29 |
| cyclopropane | 0.5584 | 0.1010 | 0.1313 | 0.0737 | 0.1502 | 0.0369 | 48.20 |
| benzene | 0.8463 | 0.1204 | 0.1517 | 0.0594 | 0.1744 | 0.0594 | 33.97 |
| thiophene | 0.7915 | 0.1200 | 0.1515 | 0.2576 | 0.2163 | 0.0953 | 14.18 |
| 2-$(CH_3)$-toluene | 1.2029 | 0.1196 | 0.1533 | 0.0798 | 0.1762 | 0.0604 | 8.03 |
| 4-$(CH_3)$-toluene | 1.1957 | 0.1262 | 0.1602 | 0.1097 | 0.1797 | 0.0565 | 4.94 |
| 3-$(CH_3)$-toluene | 1.2096 | 0.1192 | 0.1532 | 0.1055 | 0.1762 | 0.0593 | 4.10 |
| toluene | 1.0185 | 0.1208 | 0.1528 | 0.1007 | 0.1756 | 0.0810 | 3.89 |
| ethylbenzene | 1.1978 | 0.1190 | 0.1529 | 0.0826 | 0.1756 | 0.0602 | 0.54 |
| 2-$(C_6H_5)$-propane | 1.3919 | 0.1158 | 0.1528 | 0.0727 | 0.1755 | 0.0611 | 0.08 |
| Oximes, X–C(Y)=N–OH[c] | | | | | | | |
| X Y | | | | | | | |
| $CH_3S(O)CH_2C(O)$ H[d] | 0.7630 | 0.1068 | 0.1430 | 0.1687 | 0.1732 | 0.1636 | 1532 |
| $CH_3SCH_2C(O)$ H[d] | 0.5965 | 0.1043 | 0.1413 | 0.1716 | 0.1715 | 0.1628 | 1530 |
| $CH_3S(O)_2CH_2C(O)$ H[d] | 1.1435 | 0.1317 | 0.1539 | 0.1696 | 0.1816 | 0.1623 | 1477 |
| $CH_3$ H | 0.7468 | 0.1119 | 0.1357 | 0.2887 | 0.1823 | 0.1723 | 1430 |
| $CH_3$ $CH_3$ | 1.1123 | 0.1070 | 0.1489 | 0.2581 | 0.1834 | 0.1734 | 1408 |
| $C_6H_5$ H | 1.1750 | 0.1154 | 0.1436 | 0.7159 | 0.1818 | 0.2090 | 1387 |
| $CH_3C(O)$ H | 1.2463 | 0.1232 | 0.1351 | 0.6694 | 0.1985 | 0.2134 | 1379 |
| Small hydrocarbons[e] | | | | | | | |
| cyclopropane | 0.5586 | 0.1009 | 0.1313 | 0.0738 | 0.1502 | 0.0369 | 1724 |
| ethene | 0.4072 | 0.0952 | 0.1442 | 0.0799 | 0.1649 | 0.0400 | 1713 |
| ethenol | 0.4657 | 0.1022 | 0.1526 | 0.2598 | 0.1652 | 0.1872 | 1661 |
| propene | 0.5755 | 0.1027 | 0.1459 | 0.1290 | 0.1668 | 0.0524 | 1635 |
| 3-aminopropene | 0.7050 | 0.1032 | 0.1461 | 0.2763 | 0.1674 | 0.0959 | 1632 |
| propen-3-ol | 0.6446 | 0.1018 | 0.1442 | 0.3211 | 0.1675 | 0.1790 | 1632 |
| 3-fluoropropene | 0.6122 | 0.0973 | 0.1423 | 0.2433 | 0.1703 | 0.0578 | 1632 |
| fluoroethene | 0.4396 | 0.0916 | 0.1438 | 0.1913 | 0.1714 | 0.0740 | 1619 |

[a] The hydrocarbons (first set) use the activation energies for deuterium exchange instead of $\Delta H_a$. [b] Ref. 33. [c] Ref. 41. [d] Ref. 36. [e] Ref. 37.

oximes and hydrocarbons. More importantly, the equations give physically meaningful interpretations. When these results are considered along with previous applications for biological, chemical and physical properties involving solute–solvent processes, the TLSER descriptors have been shown to be very useful for a wide range of applications.

## Acknowledgements

L. Y. W. gratefully acknowledges the support of Dr. Adeny Schmidt, Dean, College of Arts and Sciences, La Sierra University and the US Army through contract number DAAD 05-92-W-0354.

## References

1 S. Gupta, *Chem. Rev.*, 1987, **87**, 1183.
2 G. N. Burkhart, *Nature* (*London*), 1935, **17**, 684.
3 L. P. Hammet, *Chem. Rev.*, 1935, **17**, 125.
4 L. P. Hammett, *J. Am. Chem. Soc.*, 1937, **59**, 125.
5 O. Exner, in *Correlation Analysis of Chemical Data*, ed. J. Shorter, Plenum Press, New York, 1988, p. 25.
6 C. Hansch, *Acc. Chem. Res.*, 1969, **B2**, 232.
7 M. J. Kamlet, R. W. Taft and J.-L. M. Abboud, *J. Am. Chem. Soc.*, 1977, **91**, 8325.
8 I. A. Koppel and V. A. Palm, *Advances in Linear Free Energy Relationships*, eds. N. B. Chapman and J. Shorter, Plenum Press, London, 1972.
9 M. J. Kamlet, R. M. Doherty, M. H. Abraham and R. W. Taft, *Quant. Struct.-Act. Relat.*, 1988, **7**, 71.
10 M. H. Abraham, G. S. Whiting, R. Fuchs and E. J. Chambers, *J. Chem. Soc., Perkin Trans. 2*, 1990, 291.
11 M. J. Kamlet, R. W. Taft, G. R. Famini and R. M. Doherty, *Acta Chem. Scand.*, 1987, **41**, 589.
12 J. P. Hickey and D. R. Passino-Reader, *Environ. Sci. Technol.*, 1991, **25**, 1753.
13 J. J. Gajewski, *J. Org. Chem.*, 1992, **57**, 5500.
14 J. G. Kirkwood, *J. Chem. Phys.*, 1934, **2**, 351; L. Onsager, *J. Am. Chem. Soc.*, 1936, **58**, 1486.
15 J. H. Hildebrand, J. M. Prausnitz and R. L. Scott, *Regular and Related Solutions*, Van Nostrand-Reinhold, Princeton, 1970.
16 B. G. Cox, G. R. Hedwig, A. J. Parker and D. W. Watts, *Aust. J. Chem.*, 1974, **27**, 477; Y. Marcus, *Pure Appl. Chem.*, 1983, **55**, 977.
17 D. F. V. Lewis, *J. Comput. Chem.*, 1987, **8**, 1084.
18 T. Brinck, J. S. Murray and P. Politzer, *Molecular Physics*, in the press.
19 J. S. Murray and P. Politzer, *J. Chem. Res.* (*S*), 1992, 110.
20 G. H. Loew, M. Poulsen, E. Kirkjian, J. Ferrell, B. S. Sudhindra and M. Rebagliati, *Environ. Health Perspect.*, 1985, **61**, 69.
21 L. Pedersen, *Environ. Health Perspect.*, 1985, **61**, 185.
22 M. Chastrette, M. Rajzmann, M. Chanon and K. F. Purcell, *J. Am. Chem. Soc.*, 1985, **107**, 1.
23 D. F. V. Lewis, in *Progress in Drug Metabollism*, eds. J. W. Bridges and L. F. Chasseaud, John Wiley, London, 1990, p. 205.
24 L. B. Kier and L. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Research Studies Press, Letchworth, 1986.
25 M. G. Ford and D. J. Livingstone, *Quant. Struct.-Adv. Relat.*, 1990, **9**, 107.
26 G. R. Famini, *Using Theoretical Descriptors in Quantitative Structure Activity Relationships* V. CRDEC-TR-085, US Army Chemical, Research, Development and Engineering Center, Aberdeen Proving Ground, MD, 1989.

27 L. Y. Wilson and G. R. Famini, *J. Med. Chem.*, 1991, **34**, 1668.

28 G. R. Famini, R. J. Kassel, J. W. King and L. Y. Wilson, *Quant. Struct.-Act. Relat.*, 1991, **10**, 344.

29 G. R. Famini, W. P. Famini, W. P. Ashman, A. P. Mickiewicz and L. Y. Wilson, *Quant. Struct.-Act. Relat.*, 1992, **11**, 162.

30 G. R. Famini, C. E. Penski and L. Y. Wilson, *J. Phys. Org. Chem.*, 1992, **5**, 395.

31 R. W. Taft, personal communication.

32 M. R. F. Siggel, M. L. Crenshaw, J. Holloway and M. S. Banna, *J. Electron Spectrosc. Relat. Phenom.*, 1990, **51**, 1.

33 M. F. Hoq and K. J. Klabunde in *Acid–Base Catal., Proc. Int. Symp.*, ed. K. Tanabe, Kodansha, Tokyo, 1988.

34 M. S. Gordon, L. P. Davis, L. W. Burggraf and R. Damrauer, *J. Am. Chem. Soc.*, 1986, **108**, 7890.

35 B. J. Smith and L. Radom, *J. Phys. Chem.*, 1991, **95**, 1059.

36 G. Bouchoux, P. Jaudon, M. Decouzon, J.-F. Gal and P.-C. Mria, *J. Phys. Org. Chem.*, 1991, **4**, 285.

37 G. D. Dahlke and S. R. Kass, *J. Am. Chem. Soc.*, 1991, **113**, 5566.

38 S. T. Graul, M. E. Schnute and R. R. Squires, *Int. J. Mass Spectrom. Ion Processes*, 1990, **96**, 181.

39 J. E. Bartmess, J. A. Scott and R. T. McIver, Jr., *J. Am. Chem. Soc.*, 1979, **101**, 6046.

40 R. Damrauer, R. Simon and M. Krempp, *J. Am. Chem. Soc.*, 1991, **113**, 4431.

41 S. G. Lias, J. E. Bartmess, J. F. Liebman, J. L. Holmes, R. D. Levin and W. G. Mallard, *J. Phys. Chem. Ref. Data, Suppl.*, 1988, **17**.

42 M. J. K. Dewar and W. Thiel, *J. Am. Chem. Soc.*, 1977, **99**, 4899.

43 J. J. P. Stewart, *Mopac Manual* FJSRL-TR-88-PP7, Frank J. Seiler Research Laboratory, US Air Force Academy, Colorado Springs, CO, 1988.

44 A. J. Hopfinger, *J. Am. Chem. Soc.*, 1980, **102**, 7126.

45 J. M. Leornard and G. R. Famini, *A User's Guide to the Molecular Modeling Analysis and Display System*, CRDEC-TR-030, US Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, 1989.

46 D. A. Belesley, E. Kuh and R. E. Welsh, *Regression Diagnostics*, Wiley, New York, 1980.

*Paper* 2/06394H
*Received* 30*th November* 1992
*Accepted* 11*th January* 1993