JOURNAL OF PHYSICAL ORGANIC CHEMISTRY, VOL. 5, 395–408 (1992)

USING THEORETICAL DESCRIPTORS IN QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS: SOME PHYSICOCHEMICAL PROPERTIES

GEORGE R. FAMINI AND CARL A. PENSKI

US Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, Maryland 21010, USA

AND

LELAND Y. WILSON

Department of Chemistry, La Sierra University, Riverside, California 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 and physicochemical properties to molecular structural features. A difficulty in this approach has been non-uniformity 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. The empirical LSER solvatochromic parameters have been replaced 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. This paper discusses the application of these descriptors to six physicochemical properties involving equilibria, kinetics and spectra. The results show good correlation and physical interpretation.

INTRODUCTION

Quantitative structure–activity relationships (QSAR) have been used extensively to correlate molecular structural features of compounds with their known biological properties (activities). QSAR assumes that there is a quantitative relationship between microscopic (molecular structure) features and a macroscopic (empirical) property (particularly biological activity) of a compound. Analogously, quantitative structure–property relationships (QSPR) apply to chemical and physical properties. Once a relationship has been 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 and 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 was given by Exner.[5]

LINEAR SOLVATION ENERGY RELATIONSHIPS

An enormous number of descriptors have been used to increase the ability to correlate biological, chemical and physical properties. Once 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) employs the empirically based solvatochromic (LSER) descriptor set and has the general form shown in the equation

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

The property is often the logarithm of a measured property (involving solute–solvent interactions) which, in turn, can be related to a free energy consistent with the LFER concept. The bulk term uses the solute molecular volume, V_1, the dipolarity terms use the solute dipolarity parameter, π^*, and the solute polarizability correction, δ, while the hydrogen bonding terms employ a solute acidity descriptor, α, and a solute basicity descriptor, β. Early work used the molar volume (V_m) while V_1 is computed. Usually not all the terms in equation (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 the 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

0894–3230/92/070395–14$12.00

Received 22 October 1991
Revised 10 March 1992

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 neighbor molecule in keeping with the Brønsted–Lowry acid–base definitions.

Thermodynamic interpretation can be inferred from work by Abraham *et al.*,[10] who correlated thermodynamic quantities for the distribution of a set of solutes between water and 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 non-polar 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 strong point of these solvatochromic (LSER) descriptors is their very successful correlation of a wide range of chemical and physical properties involving solute–solvent interactions in addition to biological activities.[11] The coefficients of the descriptors in the correlation equation also can provide an insight into the nature of the solute–solvent interactions as typified by the discussion in the previous paragraphs.

However, the LSER descriptors are limited in their ability to make *a priori* predictions because they are empirical. Although there are tables of LSER parameters and predictive relationships to help in their estimation, LSER values for complex molecules are not as easily found. Attempts to correlate computationally derived structural and electronic descriptors with the solvatochromic parameters have met with moderate degrees of success.[12]

THEORETICAL LINEAR SOLVATION ENERGY RELATIONSHIPS

In the past, theoretical chemistry has been used to provide descriptors for QSAR.[13–16] Ford and Livingstone[17] pointed out advantages of computationally derived descriptors over extrathermodynamically derived descriptors such as π and δ. These descriptors are not restricted to closely related compounds, as is often the case with group theoretical, topological and other variables. Also, they describe clearly defined molecular properties, making the interpretation of QSAR equations more straightforward, and they may be useful for prediction of biological activity. Further, 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 structures a new, theoretical set of parameters of correlating a wide variety of properties has been developed.[18,19] Termed the theoretical linear solvation energy relationship (TLSER) descriptors, these parameters are determined solely from computational methods permitting *a priori* prediction of properties. The TLSER descriptors were developed so as to give optimum correlation with the LSER descriptors, to give TLSER equations with correlation coefficients, R, and standard deviations, SD, close in value to those for LSER, and to be as widely applicable to solute–solvent interactions as the LSER set.

The TLSER bulk/steric term is described by the molecular van der Waals volume, V_{mc}, in cubic ångstroms. The dipolarity/polarizability term uses the polarizability index, π_1, 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. Aromatics rank high and alkanes low on the scale.

The hydrogen bond acceptor basicity (HBAB) is composed of covalent, ε_b, and electrostatic, q_-, basicity terms. Analogously, the hydrogen bond donor acidity (HBDA) is made up of covalent, ε_a, and electrostatic, q_+, acidity terms. Increasing ε_a and ε_b values indicate decreasing acidity and basicity, respectively. The covalent HBAB parameter, ε_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 of presentation and comparison of coefficients; the units are hectoelectron volts (heV). Analogously, the covalent HBDA parameter, ε_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 are, 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 (in acu).

The generalized TLSER equation is

$$SSP = SSP_0 + aV_{mc} + b\pi_I + c\varepsilon_b + dq_- + e\varepsilon_a + fq_+ \quad (2)$$

where SSP represents a solute–solvent interaction property; this is generally taken as the logarithm of a measured quantity. For a given property and set of compounds, the coefficients, SSP_0 and a–f are determined using multilinear regression analysis to fit the data.

The TLSER descriptors gave good correlations and physically reasonable interpretations for a set of five non-specific toxicities and a wide range of representative compounds.[19] The results compared very favorably with the LSER correlation standard deviations, SD, correlation coefficients, R, and number in the data set, N; the coefficients had the same signs and the same magnitudes, indicating qualitative agreement in the physical interpretation between the two methodologies.

The TLSER descriptors also correlate well with the minimum blocking concentrations of a set of local anesthetics[20] and with the molecular transform, a topological descriptor, which had been applied to this same set of compounds by Kier and Hall.[21] Good correlation also resulted for the opiate receptor activity of some fentanyl-like compounds and suggested an insight into the nature of the receptor sites.[22]

PROCEDURE

In this study the TLSER descriptors were applied to include six physiochemical properties (Table 1) and compound sets (Tables 5–10) that involve solute–solvent interactions. Included are charcoal adsorption,[23] high-performance liquid chromatographic (HPLC) retention indices,[24] octanol–water partition coefficients,[25] rate constants for the hydrolysis of organophosphonothiolates,[26] acidities in water (pK_a)[27,28] and electronic absorption of a pyridinium ylide.[29]

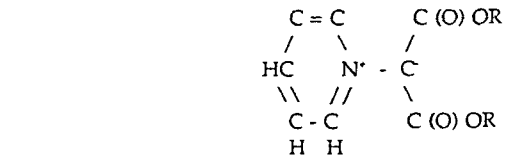
Molecular geometries were optimized and TLSER descriptors were calculated using the MNDO algorithm contained in MOPAC.[30,31] The molecular volume for

Table 1. Physiochemical properties used in this study

Property (type of process)	Symbol	Units	Reference data
Charcoal absorption	A	$l\ g^{-1}$	19
(Distribution equilibrium)	Solute (aq.) → solute (charcoal)		
HPLC retention index	K_x^a	None	20
(Distribution equilibrium)	Solute (methanol–water) → solute (column)		
Octanol–water partition coefficient	K_{ow}	None	21
(Distribution equilibrium)	Solute (aq.) → solute (octanol)		
Rate constant for hydrolysis of phosphonothiolates[b]	k_{OH}	$l\ mol^{-1}\ min^{-1}$	22
(Chemical reaction)	$RP(OR')(O)(SOR'') + OH^- \rightarrow$ products		
Acid equilibrium constant	K_a	None	23, 24
(Dissociation equilibrium)	$HA\ (aq.) + H_2O \rightarrow H_3O^+(aq.) + A^-(aq.)$		
Electronic absorption of pyridinium ylides[c]	EA	cm^{-1}	25
(Electronic transition)	e^-(ylide–solvent, Gr) → e^-(ylide–solvent, Ex)[d]		

[a] x is the percentage of methanol in water.

[b]

[c]


[d] Gr = ground state and Ex = excited state for ylide–solvent complex.

Table 2. TLSER correlations: $SSP = SSP_0 + aV_{mc}/100 + b\pi_I + c\varepsilon_b + dq_- + e\varepsilon_a + fq_+$

SSP[a]	Symbol	SSP_0	a	b	c	d	e	f	N	R	SD
Charcoal absorption, A	Log A	−6·68	2·46	5·78	n/s[b]	−4·66	n/s	n/s	33	0·955	0·24
	(t-statistic)	(−5·08)	(14·2)	(4·53)	n/s	(−4·41)	n/s	n/s			
HPCL retention index, K_{50}	LogK_{50}	−0·459	1·83	n/s	n/s	−3·05	n/s	−5·52	21	0·980	0·12
	(t-statistic	(−2·71)	(11·8)	n/s	n/s	(−10·9)	n/s	(−2·96)			
Octanol–water partition coefficient, K_{ow}	LogK_{ow}	n/s	3·14	n/s	n/s	−5·92	n/s	n/s	67	0·974	0·45
	(t-statistic)	n/s	(28·7)	n/s	n/s	(−15·5)	n/s	n/s			
Rate constant for phosphonothiolate hydrolysis, K_{OH}	LogK_{OH}	7·09	n/s	−13·1	n/s	n/s	−6·11	n/s	35	0·990	0·12
	(t-statistic)	(13·6)	n/s	(−2·98)	n/s	n/s	(−40·2)	n/s			
Acid equilibrium constant, K_a	pK_a	n/s	n/s	n/s	n/s	n/s	127	−41·1	42	0·941	2·82
	($t$statistic)	n/s	n/s	n/s	n/s	n/s	(11·3)	(−6·69)			
Electronic absorption of pyridinium ylides, EA	EA[c]	26420	n/s	−37617	n/s	1945	n/s	7613	22	0·972	239
	(t-statistic)	(36·0)	n/s	(−6·24)	n/s	(3·01)	n/s	(8·36)			

[a] SSP = solute–solvent interaction property.
[b] n/s, Not significant at the 0·95 level.
[c] The logarithm is not used as it is an energy. The logarithm of the other properties can be related to a free energy.

Table 3. Compounds with TSLER descriptors[a]

No.	Compound	$V_{mc}/100$	π_I	ε_b	q_-	ε_a	q_+	Abbreviation[b]
1	Propane	0·653	0·103	0·148	0·035	0·157	0·005	(o)
2	Butane	0·836	0·097	0·177	0·020	0·156	0·004	(o)
3	Pentane	1·002	0·100	0·163	0·081	0·156	0·045	(o)
4	2,2-Dimethylpropane	1·000	0·099	0·176	0·126	0·157	0·007	(o)
5	Hexane	1·200	0·099	0·175	0·022	0·155	0·045	(o)
6	Cyclopentane	0·895	0·102	0·175	0·017	0·154	0·009	(o)
7	Cyclohexane	1·056	0·106	0·167	0·011	0·154	0·006	(h, o)
8	Dichloromethane	0·603	0·104	0·179	0·161	0·123	0·056	(e, o, h)
9	Trichloromethane	0·753	0·113	0·184	0·112	0·115	0·088	(e, h)
10	Tetrachloromethane	0·916	0·116	0·187	0·070	0·109	0·000	(e, h, o)
12	Trichloroethene	0·862	0·116	0·161	0·072	0·117	0·109	(e, o)
13	Tetrachloroethene	1·010	0·121	0·162	0·024	0·114	0·000	(o)
14	1,1,1-Trichloroethane	0·941	0·111	0·182	0·117	0·116	0·036	(o)
15	1,2-Dichloroethane	0·773	0·106	0·179	0·185	0·121	0·049	(e)
16	1-Chloropropane	0·812	0·101	0·175	0·217	0·131	0·030	(o)
17	1,2-Dichloropropane	0·773	0·106	0·179	0·185	0·121	0·049	(c)
18	1,3-Dichloropropane	0·963	0·105	0·177	0·199	0·126	0·044	(c)
19	1-Chlorobutane	0·930	0·108	0·175	0·216	0·131	0·031	(o)
20	Methanol	0·365	0·086	0·169	0·329	0·162	0·193	(e, o)
21	Ethanol	0·542	0·093	0·167	0·324	0·157	0·180	(c, e, p, o)
22	Propan-1-ol	0·713	0·097	0·167	0·325	0·156	0·180	(c, p, o)
23	Propan-2-ol	0·721	0·096	0·166	0·320	0·155	0·178	(c, e, p, o)
24	Prop-2-en-1-ol	0·648	0·101	0·155	0·324	0·132	0·181	(c)
25	Butan-1-ol	0·898	0·098	0·167	0·325	0·155	0·180	(c, e, h, p, o)
26	Butan-2-ol	0·897	0·098	0·166	0·322	0·154	0·177	(o)
27	2-methylpropan-1-ol	0·894	0·098	0·167	0·324	0·154	0·181	(c, e, p, o)
28	2-methylpropan-2-ol	0·891	0·098	0·166	0·318	0·156	0·177	(c, o)
29	Pentan-1-ol	1·074	0·100	0·167	0·325	0·154	0·180	(c, e, h, o)
30	Pentan-3-ol	1·068	0·100	0·166	0·323	0·153	0·179	(o)
31	2,2-Dimethylpropanol	1·070	0·099	0·167	0·325	0·154	0·182	(o)
32	2-Methylbutan-2-ol	1·065	0·100	0·166	0·322	0·154	0·177	(o)
33	3-Methylbutan-2-ol	1·080	0·099	0·166	0·324	0·152	0·179	(o)

(*continued*)

Table 3. *Continued*

No.	Compound	$V_{mc}/100$	π_1	ε_b	q_-	ε_a	q_+	Abbreviation[b]
34	Hexan-1-ol	1·211	0·104	0·167	0·325	0·154	0·180	(c, e, h)
35	Cyclohexanol	1·122	0·107	0·164	0·322	0·152	0·180	(c, e, h)
36	2-Ethylbutan-1-ol	1·254	0·100	0·166	0·323	0·153	0·179	(c)
37	3,3-Dimethylbutan-2-ol	1·252	0·100	0·166	0·324	0·152	0·180	(o)
38	Octan-1-ol	1·591	0·103	0·167	0·325	0·153	0·160	(e)
39	2-Ethylhexan-1-ol	1·617	0·101	0·165	0·323	0·152	0·179	(c)
40	Propane-1,2-diol	0·793	0·095	0·167	0·332	0·153	0·185	(p)
41	Propane-1,3-diol	0·779	0·097	0·165	0·315	0·154	0·180	(p)
42	Propenal	0·566	0·106	0·158	0·301	0·122	0·061	(c)
43	Propanal	0·651	0·097	0·163	0·286	0·130	0·024	(c, o)
44	Butanal	0·824	0·099	0·162	0·289	0·130	0·024	(c, o)
45	Pentanal	1·003	0·101	0·162	0·289	0·130	0·033	(c)
46	Propanone	0·639	0·098	0·162	0·287	0·129	0·023	(c, e, p, o)
47	Butan-2-one	0·810	0·101	0·161	0·287	0·129	0·022	(c, e)
48	Pentan-2-one	1·001	0·100	0·161	0·284	0·129	0·023	(c, o)
49	4-Methylpentan-2-one	1·174	0·101	0·161	0·286	0·129	0·029	(c)
50	Hexan-2-one	1·171	0·102	0·161	0·284	0·129	0·023	(c, o)
51	Cyclohexanone	1·044	0·107	0·160	0·293	0·129	0·032	(p, o)
52	5-Methylhexan-2-one	1·360	0·101	0·161	0·287	0·129	0·033	(c)
53	Pentane-2-4-dione	0·993	0·102	0·162	0·281	0·126	0·041	(p)
54	1,1-Dichloropropanone	0·948	0·108	0·168	0·224	0·116	0·069	(p)
55	Methoxymethane	0·549	0·094	0·165	0·352	0·158	0·013	(o)
56	Ethoxyethane	0·897	0·100	0·163	0·342	0·154	0·007	(c, h, o)
57	Propoxypropane	1·267	0·101	0·163	0·345	0·153	0·018	(c)
58	2-Isopropoxypropane	1·262	0·101	0·161	0·355	0·152	0·017	(c)
59	Tetrahydrofuran	0·789	0·103	0·162	0·327	0·153	0·022	(o)
60	Methanoic acid	0·350	0·088	0·172	0·370	0·132	0·216	(p)
61	Ethanoic acid	0·529	0·095	0·170	0·370	0·130	0·220	(p)
62	Propanoic acid	0·703	0·098	0·170	0·370	0·131	0·220	(p)
63	Propenoic acid	0·628	0·105	0·162	0·378	0·121	0·215	(p)
64	Butanoic acid	0·879	0·100	0·169	0·370	0·131	0·220	(p)
65	3-Methylpropanoic acid	0·884	0·099	0·169	0·370	0·131	0·220	(p)
66	Pentanoic acid	1·058	0·101	0·169	0·370	0·131	0·220	(p)
67	3,3-Dimethylpropanoic acid	1·067	0·100	0·168	0·368	0·132	0·217	(p)
68	Hexanoic acid	1·252	0·100	0·169	0·370	0·131	0·220	(p)
69	4-Methylpentanoic acid	1·244	0·101	0·169	0·370	0·131	0·220	(p)
70	Heptanoic acid	1·426	0·102	0·169	0·370	0·131	0·220	(p)
71	Octanoic acid	1·599	0·102	0·169	0·370	0·131	0·220	(p)
72	Ethanedioic acid	0·583	0·100	0·173	0·320	0·124	0·230	(p)
73	Propane-1,3-dioic acid	0·760	0·101	0·172	0·300	0·126	0·224	(p)
74	Butane-1,4-dioic acid	0·928	0·103	0·170	0·361	0·128	0·221	(p)
75	Pentane-1,5-dioic acid	1·114	1·102	1·170	0·362	0·129	0·220	(p)
76	Hexane-1,6-dioic acid	1·298	0·103	0·170	0·362	0·129	0·218	(p)
77	Heptane-1,7-dioic acid	1·469	0·103	0·168	0·364	0·130	0·218	(p)
78	2-Heptylpropane-1,3-dioic acid	2·991	0·116	0·151	0·218	0·124	0·196	(p)
79	Methyl ethanoate	0·708	0·101	0·169	0·357	0·131	0·027	(c, e, o)
80	Ethyl ethanoate	0·889	0·102	0·169	0·357	0·131	0·026	(c, e, h, o)
81	Propyl ethanoate	1·061	0·104	0·168	0·357	0·131	0·026	(c)
82	Isopropyl ethanoate	1·074	0·102	0·168	0·340	0·132	0·027	(c)
83	Butyl ethanoate	1·210	0·106	0·168	0·357	0·131	0·026	(c)
84	Isobutyl ethanoate	1·230	0·104	0·168	0·357	0·131	0·027	(c)
85	Ethenyl ethanoate	0·821	0·110	0·152	0·349	0·127	0·083	(c)
86	Pentyl ethanoate	1·415	0·104	0·168	0·357	0·131	0·026	(c)
87	Ethyl propanoate	1·073	0·102	0·168	0·357	0·132	0·035	(o)
88	Propylamine	0·785	0·097	0·160	0·274	0·156	0·094	(p)
89	Trimethylamine	0·782	0·100	0·150	0·435	0·147	−0·036	(o)
90	Ethyldimethylamine	0·972	0·100	0·149	0·451	0·150	0·006	(o)

(*continued*)

Table 3. *Continued*

No.	Compound	$V_{mc}/100$	π_1	ε_b	q_-	ε_a	q_+	Abbreviation[b]
91	Triethylamine	1·333	0·101	0·149	0·432	0·148	0·006	(o)
92	Tripropylamine	1·848	0·103	0·148	0·436	0·148	0·101	(o)
93	*N*-Methylpiperidine	1·200	0·107	0·150	0·413	0·150	0·012	(o)
94	Acetonitrile	0·451	0·094	0·182	0·115	0·138	0·021	(e, o)
95	Propionitrile	0·632	0·097	0·180	0·106	0·138	0·029	(o)
96	Nitromethane	0·471	0·110	0·170	0·335	0·118	0·050	(o)
97	Propane-1-thiol	0·847	0·091	0·152	0·090	0·140	0·017	(p)
98	Sulfhydrylpropanone	0·841	0·092	0·155	0·284	0·127	0·042	(p)
99	2-Aminopropanethiol	0·970	0·094	0·153	0·269	0·139	0·096	(p)
100	Dimethyl sulfoxide	0·758	0·100	0·152	0·720	0·127	0·053	(o)
101	Benzene	0·846	0·120	0·148	0·059	0·159	0·059	(e, h)
102	Ethylbenzene	1·145	0·124	0·147	0·088	0·124	0·060	(h)
103	Isopropylbenzene	1·381	0·117	0·147	0·074	0·125	0·060	(h)
104	Propylbenzene	1·306	0·123	0·147	0·088	0·124	0·060	(h)
105	Bromobenzene	1·054	0·128	0·150	0·085	0·121	0·074	(o)
106	Chlorobenzene	0·994	0·124	0·151	0·112	0·121	0·078	(h, o)
107	1,4-Dichlorobenzene	1·146	0·127	0·153	0·100	0·116	0·085	(h)
108	Toluene	1·000	0·123	0·147	0·101	0·124	0·081	(e, h, o)
109	2-Methyltoluene	1·232	0·117	0·147	0·106	0·124	0·081	
110	4-methyltoluene	1·216	0·119	0·146	0·095	0·123	0·058	(h)
111	(Hydroxylmethyl)benzene	1·082	0·120	0·147	0·325	0·124	0·183	(p)
112	Phenol	0·892	0·126	0·143	0·248	0·124	0·193	(h, p)
113	3-Methylphenol	1·099	0·121	0·143	0·250	0·124	0·192	(p)
114	4-Methylphenol	1·086	0·123	0·143	0·247	0·124	0·193	(h)
115	2,4,6-Tri(*tert*-butyl)phenol	3·007	0·115	0·145	0·277	0·122	0·188	(p)
116	4-Methyl-2,6-di(*tert*-butyl)phenol	2·486	0·116	0·145	0·279	0·122	0·057	(p)
117	4-Chlorophenol	1·060	0·127	0·146	0·243	0·120	0·197	(p)
118	Benzaldehyde	1·094	0·124	0·147	0·292	0·117	0·194	(p, o)
119	Acetophenone	1·191	0·120	0·151	0·259	0·121	0·065	(o)
120	Propiophenone	1·378	0·118	0·151	0·281	0·121	0·064	(o)
121	Methoxybenzene	1·090	0·124	0·143	0·286	0·125	0·075	(e, o)
122	Ethoxybenzene	1·289	0·120	0·143	0·282	0·125	0·074	(o)
123	Propoxybenzene	1·472	0·118	0·143	0·283	0·125	0·075	(o)
124	1,2-Dimethoxybenzene	1·340	0·123	0·145	0·285	0·122	0·072	(o)
125	Ethyl benzoate	1·462	0·118	0·152	0·354	0·120	0·067	(o)
126	*N,N*-Dimethylaniline	1·314	0·125	0·137	0·415	0·126	0·060	(o)
127	*N,N*-Diethylaniline	1·658	0·119	0·147	0·387	0·124	0·071	(o)
128	4-(Dimethylamino)toluene	1·503	0·125	0·137	0·421	0·126	0·057	(o)
129	Nitrobenzene	1·017	0·131	0·158	0·342	0·110	0·095	(h)
130	Pyridine	0·791	0·122	0·151	0·230	0·122	0·084	(e)
131	Benzonitrile	0·997	0·128	0·153	0·087	0·117	0·074	(o)
132	2-Naphthol	1·337	0·141	0·139	0·248	0·118	0·194	(p)
133	$HF_2C(NO_2)$	0·537	0·104	0·175	0·304	0·112	0·114	(p)
134	$H_2ClC(NO_2)$	0·620	0·116	0·173	0·312	0·110	0·082	(p)
135	$HCl_2C(NO_2)$	0·773	0·121	0·176	0·307	0·107	0·114	(p)
136	$CH_3CS(NO_2)$	0·719	0·102	0·145	0·289	0·119	0·168	(p)
137	$F_2PO(OH)$	0·719	0·085	0·171	0·484	0·063	0·000	(p)
138	Dinaphthyl-PO(OH)	2·897	0·154	0·140	0·643	0·108	0·235	(p)
139	Di(2-ethylhexyl)-PO(OH)	3·277	0·113	0·152	0·678	0·112	0·215	(p)
140	$CH_3OPO(OH)_2$	0·758	0·102	0·168	0·636	0·114	0·237	(p)
141	$F_3CPO(OH)_2$	0·791	0·101	0·172	0·613	0·096	0·241	(p)
142	$HCl_2CPO(OH)_2$	0·994	0·117	0·166	0·625	0·104	0·236	(p)
143	$CH_3CH_2PO(OH)$	0·872	0·106	0·161	0·686	0·113	0·223	(p)
144	Uric acid	1·413	0·136	0·141	0·251	0·117	0·222	(p)
145	Purine	1·313	0·134	0·137	0·270	0·122	0·214	(p)

[a] See Table 4 for compounds used in rate of hydrolysis of phosphonothiolates.

[b] c = Charcoal adsorption; h = HPLC retention; o = octanol–water distribution; p = acid strength; e = electronic absorption.

the optimized geometry was determined using the algorithm of Hopfinger.[37] The in-house developed molecular modeling package MMADS, was used to construct and view all molecular structures.[33] Multilinear regression analysis [using MINITAB (Minitab, State College, PA, USA) or MYSTAT (Systat, Evanston, IL, USA)] was used to obtain the coefficients in the correlation equation.

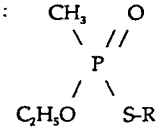
RESULTS

The correlation equations are summarized in Table 2. The compounds used and their TLSER descriptors are listed in Tables 3 and 4. Tables 5–10 list the compounds used together with the residuals for the properties. Table 11 contains the variance inflation factors (VIF) for the descriptor coefficients and the *F*-statistic values

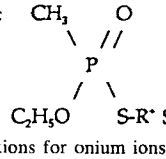
Table 4. Phosphonothiolates with descriptors

Type[a]	No.	R[a]	V_{mc}	π_1	ε_b	q_-	ε_a	q_+
Type I	P1	CH_3CH_2	1·521	0·117	0·156	0·624	0·105	0·040
	P2	$CH_3(CH_2)_2$	1·699	0·117	0·155	0·626	0·105	0·045
	P3	$CH_3(CH_2)_3$	1·892	0·115	0·155	0·597	0·104	0·048
	P4	$CH_3(CH_2)_4$	2·060	0·116	0·155	0·597	0·104	0·050
	P5	$CH_3(CH_2)_5$	2·237	0·115	0·155	0·598	0·104	0·049
	P6	$CH_3(CH_2)_6$	2·427	0·113	0·155	0·600	0·104	0·047
	P7	$CH_3(CH_2)_7$	2·602	0·112	0·155	0·626	0·105	0·046
	P8	$CH_3(CH_2)_8$	2·769	0·113	0·155	0·598	0·104	0·048
	P9	$(CH_3)_3C(CH_2)$	2·058	0·114	0·155	0·623	0·105	0·046
	P10	$(CH_3)_3C(CH_2)_2$	2·229	0·114	0·155	0·595	0·104	0·047
	P11	$(CH_3)_3C(CH_2)_3$	2·396	0·114	0·156	0·626	0·105	0·045
	P12	$(CH_3)_3C(CH_2)_4$	2·599	0·114	0·155	0·594	0·104	0·047
	P13	$(CH_3)_3C(CH_2)_5$	2·787	0·111	0·155	0·628	0·105	0·046
	P14	$(CH_3)_3C(CH_2)_6$	2·964	0·112	0·155	0·602	0·104	0·046
	P15	$(CH_3)_2CH(CH_2)$	1·878	0·115	0·155	0·624	0·105	0·047
	P16	$(CH_3)_2CH(CH_2)_2$	2·059	0·115	0·155	0·599	0·104	0·048
	P17	$(CH_3)_2CH(CH_2)_3$	2·234	0·113	0·155	0·628	0·105	0·046
	P18	$(CH_3)_2CH(CH_2)_4$	2·421	0·114	0·155	0·594	0·104	0·048
	P19	$(CH_3)_2CH(CH_2)_6$	2·782	0·111	0·155	0·628	0·105	0·046
	P20	$C_6H_5(CH_2)$	2·066	0·129	0·149	0·590	0·104	0·068
	P21	$C_6H_5(CH_2)_2$	2·228	0·128	0·148	0·592	0·104	0·062
	P22	$C_6H_5(CH_2)_3$	2·414	0·126	0·148	0·593	0·104	0·061
	P23	$C_6H_5(CH_2)_4$	2·594	0·124	0·148	0·593	0·104	0·061
	P24	$CH_3S(CH_2)_2$	1·903	0·114	0·152	0·586	0·104	0·048
	P25	$CH_3CH_2S(CH_2)_2$	2·079	0·114	0·151	0·591	0·104	0·049
	P26	$CH_3(CH_2)_2S(CH_2)_2$	2·434	0·114	0·151	0·624	0·105	0·046
	P27	$CH_3(CH_2)_5S(CH_2)_2$	2·788	0·111	0·151	0·629	0·104	0·048
	P28	$CH_3(CH_2)_7S(CH_2)_2$	3·138	0·111	0·151	0·624	0·105	0·046
	P29	$CH_3(CH_2)_9S(CH_2)_2$	3·493	0·112	0·150	0·588	0·104	0·056
Type II	P30	$CH_2CH_2S(CH_3)_2$	2·105	0·115	0·185	0·558	0·069	0·094
	P31	$CH_2CH_2S(CH_3)(C_2H_5)$	2·285	0·116	0·185	0·565	0·070	0·094
	P32	$CH_2CH_2S(CH_3)(C_4H_9)$	2·639	0·115	0·185	0·570	0·070	0·093
	P33	$CH_2CH_2S(CH_3)(C_6H_{13})$	2·980	0·115	0·184	0·568	0·071	0·094
	P34	$CH_2CH_2S(CH_3)(C_8H_{17})$	3·341	0·114	0·184	0·577	0·071	0·093
	P35	$CH_2CH_2S(CH_2)(C_{10}H_{21})$	3·704	0·113	0·185	0·571	0·071	0·093

[a] Type I:



Type II:



Calculations for onium ions. Rate data for the salts.

[the VIF is defined as $1/(1-R^2)$, where R is the correlation coefficient of one independent variable against the others; 'large' values imply strong correlation[34]]. Table 12 contains LSER correlation equations for purposes of comparison with the TLSER equations in Table 2.

Results were chosen on the basis of the t-statistic of the descriptors, the correlation coefficients, R, the standard deviations, SD, and the presence of outliers. Only terms significant at the 0·95 confidence level or higher were retained in the correlation equation. Data points with residuals greater than three standard deviations were classified as outliers.

DISCUSSION

Examination of the results in Table 2 shows the general trends and good correlations ($0{\cdot}941 < R < 0{\cdot}990$) generated by the TLSER descriptors. The three distribution equilibria have the same signs (+) for the volume and the formal charges (−). This is reasonable since the processes are similar. The volume is expected to be involved since it is a measure of the energy to form a cavity in the solvent while the positive sign indicates that larger molecules tend to be in the less polar phase. The latter fact is consistent with greater dispersive forces on larger, less polar molecules. The negative sign on the charges is reasonable because greater charge would increase charge–dipole interaction, increasing solubility in the more polar solvent. The last three properties in Table 2 depend only on electronic descriptors; in fact, each involves the HBDA. This is reasonable since these processes occur in a homogeneous phase so that a cavity does not have to be formed.

Comparison of Table 2 with Table 12 shows the qualitative agreement between the LSER and TLSER equations. The TLSER sample sets are subsets of the those used for the LSER work. For the distribution equilibria the volume term and HBAB terms have the same sign and magnitude for each set while the dipolarity/polarizability term was significant for LSER but not the TLSER. The HBDA term was only significant (barely so) for the JHPLC retention index in the TLSER case. The R and SD values are better for the LSER equations. For the electronic absorption peak for the pyridinium ylide, LSER and TLSER both indicate

Table 5. Charcoal absorption, A

No.[a]	Compound[a]	Log A Calc.	Log A Exp.	Log A Residual
17	1,2-Dichloropropane	0·48	0·57	0·09
18	1,3-Dichloropropane	0·82	1·16	0·34
21	Ethanol	−1·49	−1·35	0·14
22	Propan-1-ol	−0·84	−0·83	0·01
23	Propan-2-ol	−0·85	−1·21	−0·36
24	Prop-2-en-1-ol	−0·77	−0·89	−0·12
25	Butan-1-ol	−0·33	−0·14	0·19
27	2-methylpropan-1-ol	−0·33	−0·37	−0·04
29	Pentan-1-ol	0·22	0·32	0·10
34	Hexan-1-ol	0·79	0·96	0·17
35	Cyclohexanol	0·76	0·17	−0·59
39	2-Ethylhexan-1-ol	1·61	2·03	0·42
42	Propenal	−0·57	−0·63	−0·06
43	Propanal	−0·82	−0·71	0·11
44	Butanal	−0·26	−0·16	0·10
45	Pentanal	0·24	0·38	0·14
46	Propanone	−0·79	−0·88	−0·09
47	Butan-2-one	−0·19	−0·34	−0·15
48	Pentan-2-one	0·23	−0·19	−0·42
49	4-Methylpentan-2-one	0·70	0·71	0·01
50	Hexan-2-one	0·77	0·64	−0·13
52	5-Methylhexan-2-one	1·15	0·72	−0·43
56	Ethoxyethane	−0·29	−0·26	0·03
57	Propoxypropane	0·67	0·84	0·17
58	2-Isopropoxypropane	0·58	0·54	−0·04
79	Methyl ethanoate	−0·77	−0·64	0·13
80	Ethyl ethanoate	−0·27	−0·11	0·16
81	Propyl ethanoate	0·27	0·52	0·25
82	Isopropyl ethanoate	0·24	0·20	−0·04
83	Butyl ethanoate	0·75	1·02	0·27
84	Isobutyl ethanoate	0·69	0·60	−0·09
85	Ethenyl ethanoate	0·05	0·11	0·06
86	Pentyl ethanoate	1·14	0·82	−0·32

[a] See Table 3 for descriptors by number and name.

Table 6. HPLC retention parameter, K_{50}[a]

No.[b]	Compound[b]	Log K_{50} Calc.	Log K_{50} Exp.	Log K_{50} Residual
7	Cyclohexane	1·47	1·51	0·04
8	Dichloromethane	0·25	0·13	−0·12
9	Trichloromethane	0·58	0·48	−0·10
10	Tetrachloromethane	1·08	0·93	−0·15
25	Butan-1-ol	−0·04	−0·05	−0·01
29	Pentan-1-ol	0·21	0·32	0·11
34	Hexan-1-ol	0·40	0·55	0·14
35	Cyclohexanol	0·29	0·24	−0·04
56	Ethoxyethane	0·14	0·06	−0·08
80	Ethyl ethanoate	0·05	−0·08	−0·12
101	Benzene	0·94	0·66	−0·27
102	Ethylbenzene	1·27	1·30	0·04
103	Isopropylbenzene	1·65	1·57	−0·08
104	Propylbenzene	1·50	1·65	0·15
106	Chlorobenzene	0·94	1·01	0·07
107	1,4-Dichlorobenzene	1·19	1·38	0·18
108	Toluene	0·98	1·00	0·02
110	4-Methyltoluene	1·35	1·35	−0·00
112	Phenol	0·19	0·02	−0·17
114	4-Methylphenol	0·47	0·35	−0·12
129	Nitrobenzene	0·19	0·47	0·28

[a] 50% methanol–water.
[b] See Table 3 for descriptors by name and number.

Table 7. Octanol–water partition coefficient, K_{ow}

No.[a]	Compound[a]	Log K_{ow} Calc.	Exp.	Residual
1	Propane	1·84	2·36	0·52
2	Butane	2·50	2·89	0·39
3	Pentane	2·67	3·39	0·72
4	2,2-Dimethylpropane	2·39	3·11	0·72
5	Hexane	3·63	3·90	0·27
6	Cyclopentane	2·71	3·00	0·29
7	Cyclohexane	3·25	3·44	0·19
8	Dichloromethane	0·94	1·15	0·21
10	Tetrachloromethane	2·46	2·83	0·37
12	Trichloroethene	2·28	2·29	0·01
13	Tetrachloroethene	3·03	2·88	−0·15
14	1,1,1-Trichlorethane	2·26	2·49	0·23
16	1-Chloropropane	1·26	2·04	0·78
19	1-Chlorobutane	1·64	2·64	1·00
20	Methanol	−0·80	−0·65	0·15
21	Ethanol	−0·22	−0·30	−0·08
22	Propan-1-ol	0·31	0·28	−0·03
23	Propan-2-ol	0·37	0·05	−0·32
25	Butan-1-ol	0·89	0·99	0·10
26	Butan-2-ol	0·91	0·61	−0·30
27	2-Methylpropan-1-ol	0·89	0·76	−0·13
28	2-Methylpropan-2-ol	0·91	0·36	−0·55
29	Pentan-1-ol	1·44	1·48	0·04
30	Pentan-3-ol	1·44	1·21	−0·23
31	2,2-Dimethylpropanol	1·43	1·34	−0·09
32	2-Methylbutan-2-ol	1·43	0·89	−0·54
33	3-Methylbutan-2-ol	1·47	1·28	−0·19
34	Hexan-1-ol	1·87	2·03	0·16
37	3,3-Dimethylbutan-2-ol	2·01	1·48	−0·53
43	Propanal	0·34	0·59	0·25
44	Butanal	0·90	0·88	−0·02
46	Propanone	0·30	−0·24	−0·54
48	2-Pentanone	1·46	0·91	−0·55
50	2-Hexanone	2·04	1·38	−0·66
51	Cyclohexanone	1·66	0·81	−0·84
55	Methoxymethane	−0·36	0·10	0·46
56	Ethoxyethane	0·79	0·89	0·10
59	Tetrahydrofuran	0·54	0·46	−0·08
79	Methyl ethanoate	0·11	0·18	0·07
80	Ethyl ethanoate	0·68	0·73	0·05
87	Ethyl propanoate	1·25	1·21	−0·04
89	Trimethylamine	−0·12	0·16	0·28
90	Ethyldimethylamine	0·38	0·70	0·32
91	Triethylamine	1·62	1·45	−0·17
92	Tripropylamine	3·22	2·79	−0·43
93	*N*-Methylpiperidine	1·32	1·30	−0·02
94	Acetonitrile	0·73	−0·34	−1·07
95	Propionitrile	1·36	0·10	−1·26
96	Nitromethane	−0·51	−0·35	0·16
100	Dimethyl sulfoxide	−1·88	−1·35	0·54
101	Benzene	2·30	2·13	−0·17
105	Bromobenzene	2·80	2·99	0·19
106	Chlorobenzene	2·46	2·84	0·38
108	Toluene	2·54	2·69	0·14
109	2-Methyltoluene	3·24	3·20	−0·04
118	Benzaldehyde	1·42	1·48	0·06
119	Acetophenone	2·20	1·58	−0·62
120	Propiophenone	2·66	2·20	−0·46
121	Methoxybenzene	1·73	2·11	0·38
122	Ethoxybenzene	2·37	2·51	0·14
123	Propoxybenzene	2·94	3·18	0·24
124	1,2-Dimethoxybenzene	2·52	2·21	−0·31
125	Ethyl benzoate	2·49	2·64	0·15
126	*N,N*-Dimethylaniline	1·66	2·28	0·62
127	*N,N*-Diethylaniline	2·91	3·31	0·40
128	4-(Dimethylamino)toluene	2·22	2·61	0·39
131	Benzonitrile	2·61	1·56	−1·05

[a] Table 3 for descriptors by number and name.

dependence on electronic descriptors. The signs differ for the π term and the magnitude of the TLSER term is an order larger. The HBDA and HBAB are significant for both the TLSER and LSER relationship and have the same signs and similar magnitudes. The R and SD values are slightly better for the TLSER case. The phosphonothiolate study is an example of an advantage of TLSER; their solvatochromatic descriptors have not been measured.

There are several contributions to the less than perfect fit and the presence of outliers. First, there is the possibility of error in the measured properties. Second, the TLSER descriptors may be inadequate on several counts. The calculations are performed on isolated gas-phase molecules; correlations with solution properties could introduce significant errors. A geometry other than the global minimum may be responsible in part or whole for the activity or property.

Further confidence in the TLSER relationships comes from comparing the SD values, 0·12–2·8, with the expected experimental error in the logarithm, 0·04–0·10. the SD is much larger than the error, suggesting that the equations are not artifacts. If x is uncertain by 0·10 (10% a realistic value considering the nature of the measurements), the uncertainty in log x is 0·043. The SD of 239 cm^{-1} for the electronic absorption is larger than the spectrometer resolution (a few cm^{-1}).

Examination of the individual equations follows and shows the physically reasonable nature of the correlations. Interpretations of the relationships in Table 2 are based on the sign, magnitude and the t-statistic of the coefficients of each descriptor.

Table 5 lists the compounds, experimental and calculated values and with residuals for absorption on charcoal from water solution. Examining the TLSER in Table 2, several observations are evident. (1) The molecular volume, V_{mc}, increases absorption on charcoal; larger molecules would tend to be excluded from the

Table 8. Phosphonothiolate hydrolysis rate Constant, k_{OH}

Type[a]	No.[a]	R[a]	$-\text{Log } k_{OH}$ Calc.	Exp.	Residual
Type I	P1	CH_3CH_2	0·87	0·82	0·05
	P2	$CH_3(CH_2)_2$	0·87	0·89	−0·02
	P3	$CH_3(CH_2)_3$	0·79	0·82	−0·03
	P4	$CH_3(CH_2)_4$	0·79	0·89	−0·10
	P5	$CH_3(CH_2)_5$	0·79	0·92	−0·13
	P6	$CH_3(CH_2)_6$	0·78	0·89	−0·11
	P7	$CH_3(CH_2)_7$	0·81	0·89	−0·08
	P8	$CH_3(CH_2)_8$	0·77	0·89	−0·12
	P9	$(CH_3)_3C(CH_2)$	0·84	1·00	−0·16
	P10	$(CH_3)_3C(CH_2)_2$	0·77	0·80	−0·03
	P11	$(CH_3)_3C(CH_2)_3$	0·83	0·89	−0·06
	P12	$(CH_3)_3C(CH_2)_4$	0·77	0·82	−0·05
	P13	$(CH_3)_3C(CH_2)_5$	0·79	0·92	−0·13
	P14	$(CH_3)_3C(CH_2)_6$	0·75	0·82	−0·07
	P15	$(CH_3)_2CH(CH_2)$	0·83	0·85	−0·02
	P16	$(CH_3)_2CH(CH_2)_2$	0·79	0·80	−0·01
	P17	$(CH_3)_2CH(CH_2)_3$	0·82	0·89	−0·07
	P18	$(CH_3)_2CH(CH_2)_4$	0·77	0·80	−0·03
	P19	$(CH_3)_2CH(CH_2)_6$	0·80	0·89	−0·09
	P20	$C_6H_5(CH_2)$	0·97	0·96	0·01
	P21	$C_6H_5(CH_2)_2$	0·96	0·85	0·11
	P22	$C_6H_5(CH_2)_3$	0·93	0·96	−0·02
	P23	$C_6H_5(CH_2)_4$	0·90	0·92	−0·02
	P24	$CH_3S(CH_2)_2$	0·73	0·54	0·19
	P25	$CH_3CH_2S(CH_2)_2$	0·74	0·55	0·19
	P26	$CH_3(CH_2)_2S(CH_2)_2$	0·79	0·51	0·28
	P27	$CH_3(CH_2)_5S(CH_2)_2$	0·74	0·55	0·19
	P28	$CH_3(CH_2)27S(CH_2)_2$	0·76	0·54	0·22
	P29	$CH_3(CH_2)_9S(CH_2)_2$	0·70	0·57	0·13
Type II	P30	$CH_2CH_2S(CH_3)_2$	−1·34	−1·23	−0·11
	P31	$CH_2CH_2S(CH_3)(C_2H_5)$	−1·31	−1·23	−0·08
	P32	$CH_2CH_2S(CH_3)(C_4H_9)$	−1·28	−1·30	0·02
	P33	$CH_2CH_2S(CH_3)(C_6H_{13})$	−1·28	−1·33	0·05
	P34	$CH_2CH_2S(CH_3)(C_8H_{17})$	−1·28	−1·40	0·12
	P35	$CH_2CH_2S(CH123)(C_{10}H_{21})$	−1·30	−1·28	−0·02

[a] See Table 4 for descriptors by number and formula.

aqueous phase. This is reasonable since larger molecules have greater dispersive forces, which seem to influence lipophilic interactions. (2) The polarizability index, π_I, increases absorption on charcoal. This also makes chemical sense since greater polarizability would permit greater interaction with π-electrons on the charcoal. (3) The electrostatic (HBAB) basicity decreases absorption on charcoal. Again this is reasonable, because greater charge would tend to increase the charge–dipole interaction with the polar water molecules. This would tend to decrease the entropy and be more exothermic and exoergic; the equilibrium would favor the solute being in water.

Table 6 lists the compounds, experimental and calculated HPLC retention indices and residuals for HPLC retention in a 50% methanol–water mixture. Examination of the TLSER in Table 2 leads to the following observations. (1) The volume term increases the retention on the column as it did for charcoal adsorption. This reasonable since the larger molecules would tend to be excluded from the polar solvent [see comments under (1) for Table 5 for charcoal adsorption]. (2) The electrostatic bonding descriptors decrease the retention. Again this makes sense because a larger charge on the solute molecule would lead to a larger charge–dipole interaction with the polar methanol–water molecules. this is expected because the HPLC column is not as polar [see comments under (3) for Table 5 for charcoal]. The relationship for a 75% methanol–water solution has the same terms and signs and has $R = 0{\cdot}981$.

Table 7 lists the compounds, experimental and calcu-

Table 9. Acidity in water, K_a

No.[a]	Compound[a]	pK_a		
		Calc.	Exp.	Residual
21	Ethanol	11·23	15·90	4·67
22	1-Propanol	11·11	16·10	4·99
27	2-Methylpropan-1-ol	10·80	13·80	3·00
40	Propane-1,2-diol	10·49	14·90	4·41
41	Propane-1,3-diol	10·85	15·10	4·25
51	Cyclohexanone	14·80	16·70	1·90
53	Pentane-2,4-dione	13·98	8·88	−5·11
60	Methanoic acid	6·33	3·75	−2·58
61	Ethanoic acid	5·89	4·76	−1·13
62	Propanoic acid	6·02	4·87	−1·15
63	Propenoic acid	4·99	4·25	−0·74
64	Butanoic acid	6·02	4·82	−1·20
65	3-Methylpropanoic acid	6·02	4·86	−1·16
66	Pentanoic acid	6·02	4·86	−1·16
67	3,3-Dimethylpropanoic acid	6·29	5·05	−1·24
68	Hexanoic acid	6·02	4·88	−1·14
69	4-Methylpentanoic acid	6·02	4·84	−1·18
70	Heptanoic acid	6·02	4·89	−1·13
71	Octanoic acid	6·02	4·89	−1·13
72	Ethanedioic acid	4·65	1·27	−3·38
73	Propane-1,3-dioic acid	5·19	2·84	−2·35
74	Butane-1,4-dioic acid	5·59	4·21	−1·38
75	Pentane-1,5-dioic acid	5·76	4·34	−1·42
76	Hexane-1,6-dioic acid	5·86	4·43	−1·43
77	Heptane-1,7-dioic acid	5·99	4·48	−1·51
78	2-Heptylpropane-1,3-dioic acid	6·28	7·45	1·17
99	2-Aminopropanethiol	12·99	10·81	−2·18
112	Phenol	6·43	9·98	3·55
113	3-Methylphenol	6·48	10·10	3·62
116	4-Methyl-2,6-di(*tert*-butyl)phenol	12·72	12·20	−0·52
117	4-Chlorophenol	5·72	9·42	3·69
132	2-Naphthol	5·62	9·57	3·95
134	$H_2ClC(NO_2)$	9·99	7·20	−2·89
135	$HCl_2C(NO_2)$	8·07	5·99	−2·08
137	$F_2PO(OH)$	7·98	1·01	−6·97
138	Dinaphthyl-PO(OH)	2·38	0·74	−1·64
139	Di(2-ethylhexyl)-PO(OH)	3·85	2·85	−1·00
140	$CH_3OPO(OH)_2$	3·09	1·54	−1·55
141	$F_3CPO(OH)_2$	−0·57	1·17	0·60
142	$HCl_2CPO(OH)_2$	1·83	1·14	−0·69
143	$CH_3CH_2PO(OH)_2$	3·59	2·45	−1·14
144	Uric acid	4·14	8·40	4·26

[a] See Table 3 for descriptors by name and number.

lated partition coefficients and residuals for the distribution between water and octanol. Examination of Table 2 leads to the following observations. (1) The bulk/steric (V_{mc}) term increases the distribution into octanol. This makes sense since the larger molecules would tend to be excluded from the more polar solvent [see comments under (1) for Table 5 for charcoal absorption]. (2) The more negative the formal charge, the lesser is the tendency to dissolve in octanol. This is reasonable since a larger charge would imply greater charge–dipole interaction with the polar water molecule (see the previous two distribution equilibria).

Table 8 lists the compounds, experimental and calculated rate constants and residuals for the hydrolysis of phosphonothiolates (for formula, see Table 1). Examination of Table 2 makes the following observations evident. (1) The rate constant depends only on the electronic descriptors; the bulk/steric term of the substituents in this set of compounds was not significant. This could imply that the mechanism involves attack near

Table 10. Electronic absorption of a pyridinium ylide,[a] *EA*

No.[b]	Compound[b]	*EA* Calc.	*EA* Exp.	*EA* Residual
8	Dichloromethane	23247	23160	−87
9	Trichloromethane	23056	23280	223
10	Tetrachloromethane	22192	21919	−282
12	Trichloroethene	23026	22500	−526
15	1,2-Dichloroethane	23165	23090	−75
20	Methanol	25293	25230	−64
21	Ethanol	24922	24970	48
23	Propan-2-ol	24786	25000	214
25	Butan-1-ol	24736	24550	−186
27	2-Methylpropan-1-ol	24741	24700	−41
29	Pentan-1-ol	24660	24350	−311
34	Hexan-1-ol	24510	24520	10
35	Cyclohexanol	24391	24600	209
38	Octan-1-ol	24395	24560	165
46	Propanone	23467	23450	−17
79	Methyl ethanoate	23520	23400	−120
80	Ethyl ethanoate	23475	23300	−175
94	Acetonitrile	23267	23750	483
101	Benzene	22470	22550	80
108	Toluene	22606	22720	114
121	Methoxybenzene	22883	23040	157
130	Pyridine	22917	23100	183

[a] See Table 1 for formula.
[b] See Table 3 for descriptors by name and number.

Table 11. More statistical parameters for Table 2 correlations

Property	VIF[a]	Descriptor	*F*-statistic
Charcoal absorption	1·2	V_{mc}	100
	1·1	π_I	
	1·1	q_-	
HPLC retention index	1·1	V_{mc}	141
	1·5	q_-	
	1·5	q_+	
Octanol–water partition coefficient	4·1	V_{mc}	598
	4·1	q_-	
Rate constant for phosphonothiolate hydrolysis	1·0	π_I	817
	1·0	ε_a	
Acid equilibrium constant	10·6	ε_a	153
	10·6	q_+	
Electronic absorption of a pyridinium ylide	1·5	π_I	104
	1·8	q_-	
	1·4	q_+	

[a] Variance inflation factor for descriptor.

the P—S bond away from the R groups, since groups with larger size near the reactive site would sterically interfere. (2) The rate decreases with increasing polarizability, implying that greater ease of distortion of the electron cloud interferes with the reaction site. If the attacking group is the hydroxide ion, this further suggests that the electrons could be repelled from the reactive site, perhaps reducing the electron density below a minimum needed. (3) The rate increases with increasing hydrogen bonding acidity. This is reasonable since greater hydrogen bonding acidity implies greater ease in forming a complex with a base such as water or hydroxide ion.

Table 9 lists the compounds, experimental and calculated pK_a values and residuals for the aqueous acidity. Examination of Table 2 leads to the following observations. (1) Only the electronic descriptors are significant (the hydrogen bonding terms); this is reasonable since the acidity involves interactions with water molecules on the electronic level. (2) The signs for these electronic (HBDA) terms show that the pK_a is modelled in the expected manner.

The standard deviation is rather high; it amounts to determining K_a within a factor of 10.[3] Consequently, the regression equation in Table 2 has limited predictive value. The set of compounds includes pK_a values from 1 to 17. Taft[35] suggested that compounds with large pK_a values be excluded. When compounds with pK_a values greater than 13 were excluded from the

Table 12. LSER correlations[a] (for comparison with some TLSER equations, Table 2): $SSP = SSP_o + AV_I/100 + B\pi^* + C\beta + D\alpha$

SSP[a]	Symbol	SSP_o	A	B	C	D	N	R	SD
Charcoal absorption, *A*	Log*A*	1·93	3·06	0·56	−3·20	n/s[b]	37[c]	0·974	0·19
HPLC retention index, K_{50}	LogK_{50}	−0·38	3·22	−0·44	−2·38	n/s	29[d]	0·997	0·04
Octonol–water partition coefficient, K_{ow}	LogK_{ow}	0·20	2·74	−0·92	−3·49	n/s	102[e]	0·989	0·17
Electronic absorption of pyridinium ylides, *EA*	*EA*	21834	n/s	1310	1001	1799	22[f]	0·967	262

[a] V_I or V_m, molar volume, used; δ term not employed. *SSP* = solute–solvent interaction property.
[b] n/s, Not significant at the 0·95 level by Student's *t*-test.
[c] Ref. 36.
[d] Ref. 37.
[e] Ref. 25.
[f] Ref. 38 (same data set as used in TLSER work).

correlation, the result was the equation

$$pK_a = (29{\cdot}4) - 155\varepsilon_b \quad -7{\cdot}33q_- + 55{\cdot}4\varepsilon_a - 14{\cdot}2q_+$$

(t-statistic):

$$(9{\cdot}33)\ (-9{\cdot}05)\ (-3{\cdot}90) \qquad (3{\cdot}72)\ (-4{\cdot}96)$$

$$N = 39;\ R = 0{\cdot}950;\ \mathrm{SD} = 1{\cdot}06;\ F = 78$$

The equation gives a better correlation, 0·950 vs 0·941, a much better standard deviation, 1·06 vs 2·82, but a poorer F-statistic, 78 vs 153. Moreover, the variance inflation factors were all 2·0 or less, indicating a small cross-correlation for this data set. The pK_a values are modelled by the electronic parameters as expected. The negative sign on q_- can be rationalized by noting that the negative atom, oxygen or nitrogen, will be accompanied by a more positive (more acidic) hydrogen in the neutral molecule.

Despite the better parameters associated with equation (3), we used the correlation equation in Table 2 because of its wider pK_a range. When only the 19 caboxylic acids in Table 9 are used, the correlations improved with $R = 0{\cdot}985$ and $\mathrm{SD} = 0{\cdot}23$. The electronic descriptors, π_I, q_-, ε_a and q_+, are significant and effect the pK_a in the expected manner.

The position of an electronic absorption peak (cm^{-1}) was examined for three pyridinium ylides in the solvents listed in Table 10; the best fit was for an ylide with two ethoxy groups (see Table 1). These are solvent descriptors whereas the first five correlations involve solute descriptors. Examination of Table 2 makes several observations evident. (1) The absorption peak position depends only on electronic descriptors; this is reasonable since it involves transitions in electronic states. (2) The polarizability index decreases the position of the absorption peak (a red shift). This implies an interaction of the π-electrons on the ylide ring with the electrons of the solvent to decrease the energy gap. (3) The electrostatic acidity (more statistically significant) and basicity increase the position of the absorption peak (a blue shift), which suggests an interaction of an acidic site on the solvent molecule with a basic site on the pyridinium ylide (maybe the negatively charged carbon next to the nitrogen on the ring) increasing the energy gap. The basicity contribution can be interpreted analogously with respect to the negative nitrogen in the pyridine ring. The relationships for the two other ylides have the same terms and signs as above and R values of 0·904 and 0·930.

Outliers

For charcoal adsorption two compounds were outliers, 2-ethylbutan-1-ol and 2-methylpropan-2-ol; however, a compound with a similar structure, 2-methylpropan-1-ol, was not an outlier. The reason for this is not apparent.

For the hydrolysis of the phosphonothiolates the only outlier was hexamethylphosphoramide. Based on its V_{mc} and q_- values, its log K_{ow} value should be much larger (1·91) than it is (0·29). This compound has been an outlier in other correlations.

For the electronic absorption of the pyridinium ylide, acetonitrile was an outlier; however, it was retained in the correlation. Deleting it brought R to 0·983 and lowered the SD to 196, but then trichloroethaene became an outlier. Omitting the latter, R became 0·989 and the SD became 153. Acetonitrile, which is small and polar, has been an outlier in other correlations. In this connection, acetonitrile and hexamethylphosphoramide are molecules with portions of high charge density; the TLSER method does not handle this well.

CONCLUSIONS

Based on the application of the TLSER descriptors to the six physicochemical properties in this paper and the activities in Refs. 19, 20 and 22, the following points can be made. The TLSER descriptors have given good to very good correlations and physically reasonable interpretations. They have been applied to a wide range of properties and compounds. In addition, interpretation of the correlation equations can suggest modes of interaction between solute and solvent molecules. Where common studies have been done the TLSER correlations compare favorably with LSER results. In most cases LSER descriptors give better correlations. It should be noted that LSER has been applied to far more properties than has the TLSER.

Consequently, it appears that these computationally based TLSER parameters can be used in the same way as the empirically based LSER parameters. While LSER parameters can be estimated for new compounds from tables and predictive relationships, an attractive aspect of TLSER descriptors is that they are obtained from computation and not experiment. This permits near *a priori* prediction of properties and correlations for compounds for which the solvatochromic parameters are not readily available, such as the phosphonothiolates studied in this work.

REFERENCES

1. S. Gupta, *Chem. Rev.* **87**, 1183 (1987).
2. G. N. Burkhardt, *Nature (London)* **17**, 684 (1935).
3. L. P. Hammett, *Chem. Rev.* **17**, 125 (1935).
4. L. P. Hammett, *J. Am. Chem. Soc.* **59**, 96 (1937).
5. O. Exner, in *Correlation Analysis of Chemical Data*, p. 25, edited by J. Shorter, Plenum Press, New York (1988).
6. C. Hansch, *Acc. Chem. Res.* **B2**, 232 (1969).
7. M. J. Kamlet, R. W. Taft and J.-L. M. Abboud, *J. Am. Chem. Soc.* **91**, 8325 (1977).
8. I. A. Koppel and V. A. Palm, *Advances in Linear Free Energy Relationships*, N. B. Chapman and J. Shorter (Eds), 203–280, Plenum Press, London (1972).

9. M. J. Kamlet, R. M. Doherty, M. H. Abraham and R. W. Taft, *Quant. Struct.-Act. Relat.* **7**, 71 (1988).
10. M. H. Abraham, G. S. Whiting, R. Fuchs and E. J. Chambers, *J. Chem. Soc., Perkin Trans.* **2**, 291 (1990).
11. M. J. Kamlet, R. W. Taft, G. R. Famini, and R. M. Doherty, *Acta Chem. Scand.* **41**, 589 (1987).
12. D. F. V. Lewis, *J. Comput. Chem.* **8**, 1084 (1987).
13. G. Bouchoux, P. Jaudon, M. Decouzon, J. F. Gal and P.-C. Maria, *J. Phys. Org. Chem.* **4**, 285 (1991).
14. L. Pedersen, *Environ. Health Perspect.* **61**, 185 (1965).
15. G. H. Loew, M. Poulsen, E. Kirkjian, J. Ferrell, B. S. Sudhindra and M. Rebagliati, *Environ. Health Perspect.* **61**, 69 (1985).
16. M. Chastrette, M. Rajzmann, M. Chanon, and K. F. Purcell, *J. Am. Chem. Soc.* **107**, 1 (1985).
17. M. G. Ford and D. J. Livingstone, *Quant. Struc.-Act. Relat.* **9**, 107 (1990).
18. G. R. Famini, *Using Theoretical Descriptors in Structure Activity Relationships: V.* CRDEC-TR-085, US Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD (1989).
19. L. Y. Wison and G. R. Famini, *J. Med. Chem.* **34**, 1668 (1991).
20. G. R. Famini, R. J. Kassel, J. W. King and L. Y. Wilson, *Quant. Struct.-Act. Relat.* **10**, 344 (1991).
21. L. B. Kier, and L. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Research Studies Press, Letchworth (1986).
22. G. R. Famini, W. P. Ashman, A. P. Mickiewicz and L. Y. Wilson, *Quant. Struct.-Act. Relat.* in press.
23. H. Abe, K. Hayashi, M. Kitigawa, M. and T. Uruhata, *Bull. Chem. Soc. Jpn.* **53**, 1199 (1980).
24. T. L. Hafkensheid and E. Tomlinson, *Int. J. Pharm.* **17**, 1 (1983).
25. M. J. Kamlet, R. W. Taft, J.-L. M. Abboud, G. R. Famini, R. M. Doherty and M. J. Abraham, *J. Pharm. Sci* **74**, 807 (1985).
26. M. I. Kabachnik, A. A. Abduvakhabov, I. I. Agabekova, A. P. Brestkin, R. I. Volkova, N. N. Godovikov, E. I. Godyna, S. S. Mikhailov, M. Ya. Mikhel'son, V. I. Rozengart, E. V. Rozengart and R. V. Sitkevich, *Russ. Chem. Rev.* **39**, 485 (1970).
27. G. Körtum, W. Vogel and K. Andrewsson, *Pure Appl. Chem.* **1**, 190 (1960).
28. E. Serjent and B. Dempsey, *Ionic Constants of Organic Acids: Aqueous Solutions*, Pergamon Press, New York (1979).
29. D. Dorohoi, D. Iancu and G. Surpateanu, *Analele stiintifico de Universitatii, Al. I. Cuza din Iasi 1b*, **27**, 59 (1981). [A Romanian Journal].
30. M. J. K. Dewar and W. Thiel, *J. Am. Chem. Soc.* **99**, 4899 (1977).
31. J. J. P. Stewart, *Mopac Manual.* FJSRL-TR-88-0007, Frank J. Seiler Research Laboratory, US Air Force Academy, Colorado Springs, CO (1988).
32. A. J. Hopfinger, *J. Am. Chem. Soc.* **102**, 7126 (1980).
33. J. M. Leonard 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).
34. D. A. Belsley, E. Kuh and R. E. Welsh, *Regression Diagnostics*, Wiley, New York (1980).
35. R. W. Taft, personal communication, 1991.
36. M. J. Kamlet, R. M. Doherty, M. H. Abraham and R. W. Taft, *Carbon*, **23**, 549 (1985).
37. D. E. Leahy, P. W. Carr, R. S. Pearlman, R. W. Taft and M. J. Kamlet, *Chromatographia*, **21**, 473 (1986).
38. G. R. Famini, C. A. Penski and L. Y. Wilson. Unpublished results.