

Correlation of the Solubilities of Gases and Vapors in Methanol and Ethanol with Their Molecular Structures

Alan R. Katritzky* and Douglas B. Tatham

Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, P.O. Box 117200, Gainesville, Florida 32611-7200

Uko Maran

Department of Chemistry, University of Tartu, 2 Jakobi Street, Tartu 51014, Estonia

Received August 25, 2000

Two four-parameter quantitative structure–property relations, with $R^2 = 0.95$ and $R^2 = 0.97$, respectively, gave good correlations for the solubilities of 87 gases and vapors in methanol and 61 in ethanol. All the descriptors used are derived solely from the structures of the molecules, making it possible to predict solubilities for unavailable or unknown compounds.

INTRODUCTION

We recently showed¹ that the solubilities of a diverse set of 406 gases and vapors in water were correlated by the five-parameter eq 1 with $R^2 = 0.942$. The descriptors of eq 1 are all derived solely from the chemical structures and do not require any additional experimental data. Thus, unlike previous treatments,^{2,3} eq 1 can be used for the prediction of the aqueous solubilities of additional structures for compounds that are unavailable or even unknown.

$$\log L^w = 2.65 + 42.37\text{HDCA}(2) + 0.65[\text{N}(\text{O}) + 2^*N(N)] - 0.16(E_{\text{HOMO}} - E_{\text{LUMO}}) + 0.12 \text{PCWT}^E + 0.82N_{\text{rings}} \quad (1)$$

In connection with our ongoing work on quantitative structure–property relationships in general,^{4,5} and our interest in solvent scales⁶ and solvent properties⁷ in particular, we plan a general treatment of solubilities. The present paper represents the first stage in this work and deals with solubilities of gases and vapors in methanol and ethanol, i.e., with gas–liquid partition coefficients. Relevant previous work in this area has been carried out by the Abraham group, who used a modified *linear solvation energy relationship* (LSER) approach that was extended to model the solubility and distribution behavior of nonelectrolyte solutes with solvent properties.⁸ For the solubilities of gases and vapors in methanol at 298 K they reported⁹ eq 2, which relates $\log L^{\text{MeOH}}$ with the excess molar refraction (R_2), the polarizability (π_2^H), the overall hydrogen bond acidity ($\Sigma\alpha_2^H$), the overall hydrogen bond basicity ($\Sigma\beta_2^H$), and the Ostwald solubility coefficient in hexadecane (L^{16}). For 93 substrates, eq 2 has $R^2 = 0.9952$. Although this correlation gives much insight into the physical basis of gas-methanol partition coefficients, it cannot easily be used for the prediction of such coefficients for unavailable compounds because it uses experimentally determined descriptors.

$$\log L^{\text{MeOH}} = -0.004 - 0.215R_2 + 1.173\pi_2^H + 3.701\Sigma\alpha_2^H + 1.432\Sigma\beta_2^H + 0.769 \log L^{16} \quad (2)$$

The Abraham group has also carried out similar work¹⁰ for the solubilities of gases and vapors in ethanol. They derived eq 3 ($R^2 = 0.9966$), which links these solubilities ($\log L^{\text{EtOH}}$) with the parameters as already described for methanol. A major limitation of the descriptors involved in eq 3 is their experimental origin.

$$\log L^{\text{EtOH}} = -0.012 - 0.206R_2 + 0.789\pi_2^H + 3.635\Sigma\alpha_2^H + 1.311\Sigma\beta_2^H + 0.853 \log L^{16} \quad (3)$$

Our aim was to derive equations with satisfactory predictive power for solubilities in methanol and ethanol using solely descriptors that can be derived from the molecular structures alone. The present paper demonstrates that this is indeed possible.

RESULTS AND DISCUSSION

Solubilities in Methanol. We treated 87 Ostwald solubility coefficients (Table 1) taken from the Abraham paper.⁹ These represent the entire data set from that paper with the following exceptions: helium, neon, argon, krypton, xenon, and radon. The noble gases were omitted because atomic species are not appropriate for the optimization methods (see Experimental Section) that we use. In particular, the above-mentioned gases are not parametrized in the frame of the AM1 parametrization that is used for calculating approximately 550 quantum chemical descriptors per compound depending on the atomic constitution of the molecule. This is approximately 2/3 of the descriptor pool.

$$\log L^{\text{MeOH}} = -1.1113 + 0.0532\pi^{\alpha} + 29.163\text{HDCA}_{\text{TMSA}^{1/2}}^{2M} + 0.4195\mu_{\text{tot}} + 0.8871 {}^0\overline{IC} \quad (4)$$

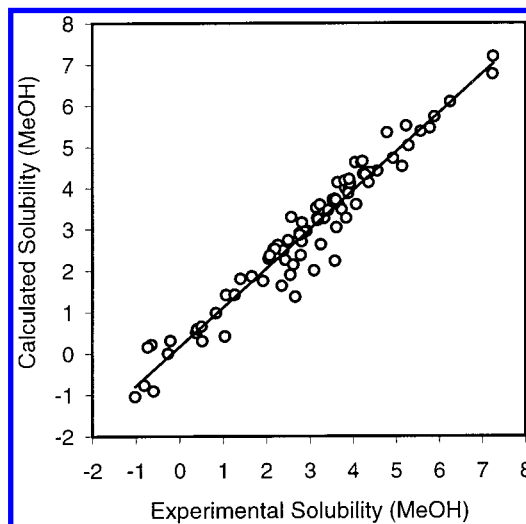
$$(R^2 = 0.9446, R_{\text{cv}}^2 = 0.9382, 87 \text{ structures})$$

* Corresponding author phone: (352)392-0554; fax: (352)392-9199; e-mail: katritzky@chem.ufl.edu.

Table 1. Data Set for Methanol Solubility Study

compound	$L^{\text{MeOH}}(\text{exptl})$	$L^{\text{MeOH}}(\text{calcd})$	difference	compound	$L^{\text{MeOH}}(\text{exptl})$	$L^{\text{MeOH}}(\text{calcd})$	difference
1,1,2,2-tetrachloro-1,2-difluoroethane	2.76	2.90	0.14	methane	-0.27	0.00	0.27
2-methylpropane	1.07	1.41	0.34	methyl butanoate	3.55	3.62	0.07
ammonia	2.26	2.61	0.35	methyl propanoate	3.17	3.26	0.09
biphenyl	5.78	5.46	-0.32	nitrogen	-0.81	-0.77	0.04
butanone	3.33	3.26	-0.07	nonane	3.23	3.59	0.36
butyraldehyde	3.18	3.22	0.04	oxygen	-0.60	-0.91	-0.31
chlorobenzene	3.67	3.59	-0.08	pentanal	3.57	3.62	0.05
dibutyl ether	3.64	4.13	0.49	1-propanol	4.36	4.13	-0.23
diethylamine	3.73	3.48	-0.25	propionaldehyde	2.79	2.85	0.06
ethane	0.38	0.50	0.12	propylbenzene	3.91	4.21	0.30
ethene	0.41	0.59	0.18	tetrachloromethane	2.66	1.37	-1.29
1-heptanol	5.57	5.38	-0.19	toluene	3.19	3.24	0.05
heptan-4-one	4.23	4.34	0.11	trifluoroethanol	4.19	4.62	0.43
hexan-2-one	3.82	3.99	0.17	2,2-dimethylpropane	1.40	1.80	0.40
hexafluoro-2-propanol	4.79	5.35	0.56	acetonitrile	2.81	2.70	-0.11
isopropylbenzene	3.80	4.17	0.37	benzene	2.79	2.37	-0.42
methyl acetate	2.92	2.96	0.04	butane	1.27	1.42	0.15
methyl pentanoate	3.91	3.99	0.08	butylbenzene	4.22	4.65	0.43
naphthalene	5.14	4.52	-0.62	carbon monoxide	-0.65	0.21	0.86
1-nonanol	6.25	6.09	-0.16	cyclopentane	1.92	1.75	-0.17
octane	2.82	3.16	0.34	diethyl ether	2.39	2.48	0.09
pentan-2-one	3.58	3.62	0.04	dioxane	3.58	2.23	-1.35
phenanthrene	7.24	6.76	-0.48	ethanthiol	2.20	2.52	0.32
propanone	2.89	2.94	0.05	ethyne	1.04	0.41	-0.63
propyl acetate	3.53	3.71	0.18	heptan-3-one	4.26	4.34	0.08
<i>tert</i> -butylbenzene	4.05	4.62	0.57	1-hexanol	5.29	5.03	-0.26
tetrahydrofuran	3.25	2.63	-0.62	hexylbenzene	5.23	5.51	0.28
triethylamine	3.84	3.27	-0.57	iodomethane	2.35	1.63	-0.72
1,1-dichloroethane	2.16	2.52	0.36	methanol	3.60	3.71	0.11
3,3-diethylpentane	3.15	3.52	0.37	methyl hexanoate	4.28	4.32	0.04
anthracene	7.25	7.19	-0.06	methyl <i>tert</i> -butyl ether	2.77	2.86	0.09
1-butanol	4.56	4.41	-0.15	nitromethane	3.42	3.46	0.04
butyl acetate	3.93	4.11	0.18	1-octanol	5.89	5.73	-0.17
carbon dioxide	0.52	0.30	-0.22	1-pentanol	4.93	4.72	-0.21
cyclohexane	2.43	2.26	-0.17	pentane	1.66	1.86	0.20
dichloromethane	2.55	1.90	-0.65	propane	0.84	0.98	0.14
dimethylamine	2.57	3.29	0.72	propionitrile	3.61	3.04	-0.57
ethanol	3.89	3.87	-0.02	sulfur dioxide	2.08	2.36	0.28
ethylbenzene	3.58	3.73	0.15	tetrafluoromethane	-0.73	0.15	0.88
heptan-2-one	4.38	4.38	0.00	trichloromethane	3.10	2.00	-1.10
heptane	2.50	2.73	0.23	water	4.07	3.60	-0.47
hexane	2.05	2.29	0.24	sulfur hexafluoride	-0.21	0.31	0.52
iodoethane	2.62	2.14	-0.48	nitrous oxide	0.51	0.65	0.14
				hydrogen	-1.02	-1.04	-0.02

The results of the CODESSA treatment (see Experimental Section) of this data set (given in Tables 1–3) produced eq 4 ($R^2 = 0.9446$). Two quantum chemical descriptors are the α polarizability (π^α) and total dipole of the molecule (μ_{tot}). The polarization of a molecule is induced by an external electric field in the medium and therefore directly reflects the interaction between the medium and an embedded molecule.¹¹ The first-order polarizability term has also been shown to be related to hydrophobicity.¹² Both these descriptors describe the Coulombic interactions between the solute and the solvent. The electrostatic descriptor *modified hydrogen donor atomic charge weighted charged surface area divided by the square root of the total molecular surface area* ($\text{HDCA}_{\text{TMSA}^{1/2}}^{2\text{M}}$) (using quantum chemically calculated charge distribution) describes the ability of the solute to act as a hydrogen bond donor. It is calculated from the sum of the partial charges on hydrogen-bonding donor (H) atom(s) that is multiplied by the square root of the surface area of this atom and then the product is divided by the square root of the total molecular surface area. The surface areas are calculated from the van der Waals radii of the atoms (overlapping spheres). The descriptor accounts only for the surface areas of hydrogen in $-\text{OH}$, $-\text{NH}$, and $-\text{SH}$ groups

**Figure 1.** Solubility in methanol (eq 4).

in the compounds, explicitly showing the hydrogen donating compounds. For the compounds that do not contain those functional groups the descriptor value is zero. It is a similar descriptor to that previously used for calculating water

Table 2. Correlation of Methanol Solubility Coefficients

	X	$\pm\Delta X$	R^2	R^2_{cv}	σ^2	F	t -test	descriptors
0	$-1.1113 \times 10^{+00}$	1.6343×10^{-01}					-6.7998	intercept
1	5.3172×10^{-02}	1.8269×10^{-03}	0.5508	0.5301	1.3819	104.21	29.1054	π^a
2	$2.9163 \times 10^{+01}$	$1.8867 \times 10^{+00}$	0.7721	0.7655	0.6920	146.64	15.4574	$\text{HDCA}_{\text{TMSA}^{1/2}}^{2M}$
3	4.1945×10^{-01}	5.7387×10^{-02}	0.9254	0.9186	0.2351	343.06	7.3091	μ_{tot}
4	8.8715×10^{-01}	1.6633×10^{-01}	0.9446	0.9382	0.1767	349.49	5.3336	${}^0\text{IC}$

Table 3. Intercorrelation of Descriptors for Methanol

	π^a	$\text{HDCA}_{\text{TMSA}^{1/2}}^{2M}$	μ_{tot}	${}^0\text{IC}$
π^a	1.0000			
$\text{HDCA}_{\text{TMSA}^{1/2}}^{2M}$	-0.1745	1.0000		
μ_{tot}	-0.1911	0.2251	1.0000	
${}^0\text{IC}$	0.1685	0.0324	0.6194	1.0000

solubility.¹ The difference is in the hydrogen donor surfaces that are accounted for. In addition to the hydrogen surface areas of the above-mentioned functional groups, the hydrogens attached to carbons that are connected directly to carbonyl and cyano groups were included for water solubilities as possible hydrogen bonding donor centers (their effectiveness, of course, is much smaller because of their smaller partial charge).

The topographical descriptor *zeroth-order average information content* (${}^0\text{IC}$) is based on the Shannon information theory¹³ and uses the zero coordination sphere. This descriptor (zeroth order) simply counts the atoms and groups them into classes corresponding to the type of atom and then gives the probability that the atom will occur in the class. In this way the descriptor accounts for the similar types of atoms (addressing the size of the compound) and the information about the diversity of the atomic constitution of the compounds.

The descriptors of eq 4 describe the solute-solvent interactions during the solvation process and are also comparable with those in the Abraham equation (eq 2). The polarizability and dipole descriptors describe the nonspecific electrostatic solute-solvent interactions and echo Abraham's π_2^H descriptor and also track the L^{16} descriptor. The hydrogen bonding descriptor, $\text{HDCA}_{\text{TMSA}^{1/2}}^{2M}$, covers the relationship between the total surface of the molecule and the hydrogen bond donating portions of alcohols, amines, and thiols, reflecting the specific solute-solvent interactions. In Abraham's paper,⁹ hydrogen bonding is described via hydrogen bond acidity and basicity. There is no direct analogy for the ${}^0\text{IC}$ term in Abraham's eq 2. However due to the direct connection between the size of the compound and its atomic constitution, this term can be loosely related to the cavity formation and dispersion interactions that belong to the solute-solvent interactions. Even if this descriptor does not directly account for the shape of the cavity of the molecule, it is an important factor in both interactions.

The predictive power of eq 4 was demonstrated by cross-validation ($R^2_{cv} = 0.9382$) and by means of an internal validation set ($R^2 = 0.9419$). The method used for the internal validation set is outlined in the Experimental Section.

Solubilities in Ethanol. We treated 61 Ostwald solubility coefficients (Table 4) taken from the Abraham paper.¹⁰

Table 4. Data Set for Ethanol Solubility

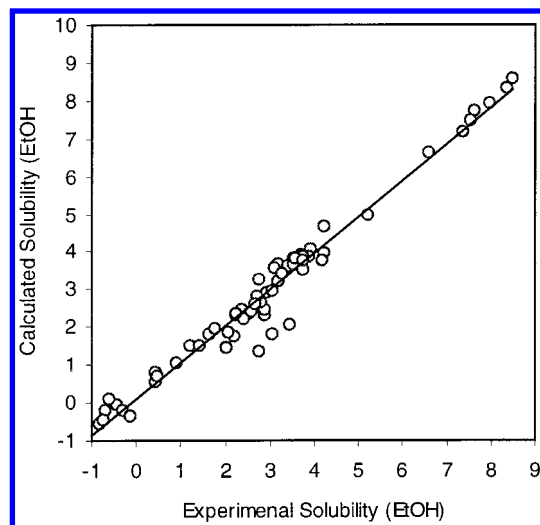
structure	$L^{\text{MeOH}}(\text{exptl})$	$L^{\text{MeOH}}(\text{calcd})$	difference	structure	$L^{\text{MeOH}}(\text{exptl})$	$L^{\text{MeOH}}(\text{calcd})$	difference
1,1-dichloroethane	2.38	2.48	0.10	iodoethane	2.59	2.40	-0.19
1,2-difluoro-1,1,2,2-tetra-chloroethane	2.87	2.32	-0.55	methane	-0.29	-0.20	0.09
1,4-dioxane	3.44	2.07	-1.37	methanol	3.40	3.61	0.21
2,2-dimethylpropane	1.61	1.81	0.20	methyl propanoate	3.04	2.97	-0.07
2-iodopropane	2.92	2.93	0.01	naphthalene	5.21	4.97	-0.24
2-methylpropane	1.20	1.49	0.29	nitrogen	-0.84	-0.54	0.30
acetonitrile	2.90	2.45	-0.45	nitromethane	3.10	3.59	0.49
ammonia	2.01	1.46	-0.55	nitrous oxide	0.45	0.70	0.25
anthracene	7.62	7.73	0.11	nonadecane	7.95	7.92	-0.03
benzene	2.78	2.65	-0.13	nonane	3.52	3.66	0.14
bromoethane	2.22	2.31	0.09	octadecane	7.54	7.49	-0.05
iodomethane	2.19	1.78	-0.41	octane	3.17	3.23	0.06
butane	1.40	1.53	0.13	oxygen	-0.62	0.12	0.74
butanone	3.16	3.65	0.49	pentane	1.78	1.96	0.18
butylbenzene	4.20	4.69	0.49	phenanthrene	7.36	7.21	-0.15
carbon dioxide	-0.42	-0.05	0.37	propan-1-ol	4.20	3.99	-0.21
carbon monoxide	-0.70	-0.22	0.48	propan-2-ol	3.86	3.89	0.03
chlorobenzene	3.68	3.93	0.25	propane	0.91	1.08	0.17
cyclohexane	2.55	2.35	-0.20	propanone	2.77	3.27	0.50
cyclopentane	2.04	1.85	-0.19	<i>p</i> -xylene	3.73	3.85	0.12
decane	3.92	4.08	0.16	pyrene	8.47	8.60	0.13
dimethylamine	2.40	2.23	-0.17	pyridine	4.18	3.76	-0.42
ethane	0.44	0.55	0.11	tetrachloromethane	2.75	1.35	-1.40
ethanol	3.73	3.51	-0.22	tetrafluoromethane	-0.72	-0.47	0.25
ethene	0.41	0.79	0.38	toluene	3.28	3.40	0.12
ethylbenzene	3.51	3.80	0.29	trichloromethane	3.06	1.83	-1.23
heptane	2.71	2.81	0.10	triethylamine	3.57	3.84	0.27
hexadecane	6.57	6.64	0.07	trimethylamine	2.67	2.60	-0.07
hexane	2.24	2.38	0.14	water	3.72	3.79	0.07
eicosane	8.34	8.35	0.01	sulfur hexafluoride	-0.12	-0.37	-0.25
				hydrogen	-1.07	-0.92	0.15

Table 5. Correlation of Ethanol Solubility Coefficients

	X	$\pm\Delta X$	R^2	R^2_{cv}	S^2	F	t -test	descriptors
0	$2.1110 \times 10^{+00}$	5.1109×10^{-01}					4.1305	intercept
1	5.2338×10^{-02}	1.6691×10^{-03}	0.7724	0.7596	1.2076	200.18	31.3563	π^a
2	6.7322×10^{-01}	6.3197×10^{-02}	0.9023	0.8883	0.5270	267.97	10.6527	μ_{tot}
3	4.9708×10^{-02}	5.1666×10^{-03}	0.9536	0.9468	0.2548	390.42	9.6209	HASA
4	2.0821×10^{-01}	4.0325×10^{-02}	0.9686	0.9640	0.1757	431.29	5.1633	ϵ_{HOMO}

Table 6. Intercorrelation of Descriptors for Ethanol

	π^a	μ_{tot}	HASA	ϵ_{HOMO}
π^a	1.0000			
μ_{tot}	-0.3363	1.0000		
HASA	-0.2160	0.2907	1.0000	
ϵ_{HOMO}	0.5031	0.0546	-0.0640	1.0000

**Figure 2.** Solubility in ethanol (eq 5).

Helium, neon, argon, krypton, xenon, and radon were omitted from our database for the same reasons as given above in the methanol study.

$$\log L^{EtOH} = 2.1110 + 0.0523\pi^a + 0.6732\mu_{tot} + 0.0497HASA + 0.2081\epsilon_{HOMO} \quad (5)$$

$$(R^2 = 0.9686, R^2_{cv} = 0.9640, 61 \text{ structures})$$

The results of the CODESSA treatment (see Experimental Section) for this data set (given in Tables 4–6) produced eq 5 ($R^2 = 0.9686$). As with the methanol correlation, two of the quantum chemical descriptors, α polarizability (π^a) and total dipole of the molecule (μ_{tot}), describe the Coulombic interactions between the solute and the solvent as already described for methanol solubility. The electrostatic descriptor hydrogen-acceptors surface area (HASA) uses the quantum chemically calculated charge distribution on atoms and accounts directly for the ability of the solute to act as a hydrogen bond acceptor. HASA is the surface area of the compound that is capable of accepting a hydrogen in an interaction with the environment. If a compound contains a hydrogen with a partial charge higher than 0.185, the descriptor will search for the N and O atoms and then sum the partial surface areas of those atoms. Generally these hydrogens are part of $-OH$ or $-NH$ functional groups. For the current data set, the different behavior of $-OH$ oxygen-linked hydrogen is identified because the $-NH$ hydrogens

have a lower partial charge. HASA is a descriptor very similar to $HDCA_{TMSA1/2}^{2M}$ in the methanol solubility model. Both descriptors account for the similar functional groups. Therefore, one can say that in HASA the hydrogen bond donor capability of compounds is represented in the hidden way and vice versa.

The remaining quantum chemical descriptor *energy of the highest occupied molecular orbital* (ϵ_{HOMO}) describes the reactivity of the molecule, which can also be loosely related to hydrogen bonding.

The descriptors in eq 5 are similar to those found in eq 4 for the correlation of methanol and are therefore comparable to those in the Abraham equation (3). Moreover, they again stress the importance of polarizability and dipole interactions as nonspecific interactions in the solubility process. They are complemented by the hydrogen bond donor and acceptor capability terms, which represent the specific interactions that solute molecules have in solubility equilibria. The major difference between eqs 4 and 5 is that the latter contains no descriptor equivalent to the lipophilicity descriptor, L .¹⁶ However, while they echo Abrahams π_2^H descriptor in eq 3, the polarizability and dipole descriptors in eq 5 also reflect lipophilicity. The hydrogen bond descriptor in eq 5, HASA, directly shows the hydrogen bond accepting ability and indirectly shows the hydrogen bond donor capabilities. This descriptor provides alternatives to hydrogen bond basicity and acidity in Abrahams eq 3. The predictive power of eq 5 was also demonstrated by cross-validation ($R^2_{cv} = 0.9640$) and an internal validation set ($R^2 = 0.9659$).

A major difference from Abrahams approach^{9,10} is that our treatment uses a large pool of descriptors and seeks to find the strongest correlation possible. Abraham applied his equation to methanol (eq 2) and ethanol (eq 3), using the same descriptors which allowed Abraham to draw conclusions about the hydrogen bonding acidity and basicity of the solvents.

$$\log L^{MeOH} = 2.2773 + 0.0477\pi^a + 0.5903\mu_{tot} + 0.0560HASA + 0.2052\epsilon_{HOMO} \quad (4a)$$

$$(R^2 = 0.9306, R^2_{cv} = 0.9219, 87 \text{ structures})$$

$$\log L^{EtOH} = -1.1993 + 0.0537\pi^a + 25.736HDCA_{TMSA1/2}^{2M} + 0.4528\mu_{tot} + 0.9178\overline{TC} \quad (5a)$$

$$(R^2 = 0.9642, R^2_{cv} = 0.9578, 61 \text{ structures})$$

Analogously, we performed a multilinear regression for ethanol using the descriptors found in our correlation for methanol (eq 4) to give eq 5a ($R^2 = 0.9642$). Similarly, a regression on methanol solubilities using the descriptors from eq 5 gave eq 4a ($R^2 = 0.9306$). In keeping with Abraham's results,^{9,10} a comparison of equations with common descrip-

tors (i.e. eq 4 with eq 5a and eq 5 with eq 4a) demonstrated coefficients for π^α greater for ethanol than for methanol, as expected given the greater polarity of ethanol.

An advantage to finding the best correlations from a large pool of descriptors is that the choice of descriptors provides information. Both eqs 4 and 5 have α polarizability and dipole moment in common. The next most important descriptors, according to *t*-test results, are $\text{HDCA}_{\text{TMSA}^{1/2}}^{2\text{M}}$ and HASA for eqs 4 and 5, respectively. $\text{HDCA}_{\text{TMSA}^{1/2}}^{2\text{M}}$ is a descriptor for the hydrogen bond donating capability which corresponds well to Abrahams observation⁹ that the hydrogen bonding acidity of the solute plays a very significant contribution. In contrast the correlation for ethanol included HASA, a descriptor for hydrogen bond accepting capability. The different hydrogen bonding descriptor results from the lower hydrogen-bond basicity of ethanol compared to methanol, again consistent with Abrahams conclusions.¹⁰

Drawing conclusions based on the coefficients of common descriptors, when the entire descriptor set is not shared, can cause problems because of collinearity. An example of this occurs with eqs 4 and 5. The average information content is collinear with the total dipole moment, and as a result the coefficients for μ_{tot} decrease when ^0TC occurs in the equation.

GENERAL CONCLUSIONS

We conclude that the type of treatment that was successful for gas–water partition coefficients¹ can be applied to other solvents as demonstrated in the present paper for methanol and ethanol. The result of the treatment is highly consistent with previous works on development of empirical descriptors for the multilinear analysis of solute–solvent interactions by Katritzky et al.,¹⁴ Koppel and Palm,¹⁵ Kamlet and Taft,¹⁶ Krygowsky and Fawcett,¹⁷ Sawin et al.,¹⁸ Mayer,¹⁹ and Dougherty.²⁰ The descriptors in our equations can be more broadly classified to account for nonspecific (polarizability and dipole interactions) and specific (hydrogen bonding) interactions. The advantages of this treatment are in the simple derivation of equations presented. Additionally, our equations both have predictive powers for the determination or the estimation of such gas–liquid partition coefficients for unavailable or unknown compounds and also throw light on the factors driving the magnitude of such parameters. Along with this, the applicability of the quantum chemical calculations in the development of descriptors is shown. All four of the descriptors for ethanol, and three of the four descriptors for methanol, have quantum chemical origins.

It should now be feasible to extend this treatment to other solvents, which could eventually provide a general treatment of solubility similar to the general treatment of solvent polarity that we previously reported.⁵

EXPERIMENTAL SECTION

The molecular structures were drawn and preoptimized by the MMX molecular mechanics method incorporated in PCMODEL.²¹ Final optimizations were performed using the AM1 parametrization²² and eigenvector following geometry optimization procedure²³ within the semiempirical quantum-chemical program MOPAC 6.0.²⁴ A gradient norm 0.01 kcal/Å was used to calculate electronic, geometric, and energetic parameters for the isolated molecules. AM1 force

calculations were used to produce thermodynamic constants. After optimization, the structures were used in the CODESSA program to calculate constitutional, topological, geometrical, thermodynamic, quantum chemical, and electrostatic descriptors.^{25,26} CODESSA produced approximately 800 to 820 descriptors for each of the structures in both sets. The best multilinear regression (BMLR) procedure^{26–28} was used to find the optimum correlation models. The BMLR procedure selected the best two-parameter regression model, the best three-parameter regression model, etc. based on the highest R^2 value in a forward stepwise regression procedure.²⁹ The correlation equations were constituted from the selected noncollinear descriptors according to the Fisher criteria. The final choices had the maximum value of the Fisher criteria and the highest value of the cross-validated correlation coefficient (R_{cv}^2) and were considered to be the best criteria for a relationship utilizing the given descriptor pool.^{28,26,30}

The initial correlations found some descriptors, such as number of fluorine atoms, number of chlorine atoms, or number of rings, which appeared to be artifacts of the relatively small data sets and do not give a general understanding of solute–solvent interactions in solubility process. Such descriptors were eliminated from the pool and new correlations were performed, providing models utilizing general descriptors.

Correlation eqs 4 and 5 were validated using internal validation sets. The full set of structures was divided into three groups, structures 1, 4, 7, etc. formed Group 1, structures 2, 5, 7, etc. formed Group 2, and structures 3, 6, 8, etc. formed Group 3. Groups 1 and 2 were then combined to form Set A, Groups 2 and 3 to form Set B, and Groups 1 and 3 to form Set C. Correlations were performed on Set A, Set B, and Set C using the descriptors from eqs 4 and 5. The partition coefficients of the descriptors for sets A, B, and C were recorded. The solubilities for Group 3 were then predicted using the descriptor partition coefficients from Set A (generated from Groups 1 and 2—see above), those of Group 1 were predicted using the partition coefficients from Set B, and Group 2 solubilities were predicted using the partition coefficients from Set C. The solubilities calculated for Groups 1, 2, and 3 were combined, and a correlation was then performed between the calculated and the experimental solubility values to give the R^2 values.

REFERENCES AND NOTES

- (1) Katritzky, A. R.; Mu, L.; Karelson, M. A QSPR Study of the Solubility of Gases and Vapors in Water. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1162–1168.
- (2) Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. Hydrogen Bonding. Part 34. The Factors that Influence the Solubility of Gases and Vapours in Water at 298 K, and a New Method for its Determination. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777–1791.
- (3) Yalkowsky, S. H.; Valvani, S. C. Solubility and Partitioning I: Solubility of Nonelectrolytes in Water. *J. Pharm. Sci.* **1980**, *69*, 912–922.
- (4) Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A. R. QSPR and QSAR Models Derived Using Large Molecular Descriptor Spaces. A Review of CODESSA Applications. *Collect. Czech. Chem. Commun.* **1999**, 1551–1571.
- (5) Katritzky, A. R.; Maran, U.; Lobanov, V. S.; Karelson, M. Structurally Diverse Quantitative Structure–Property Relationship Correlations of Technologically Relevant Physical Properties. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1–18.
- (6) Katritzky, A. R.; Tamm, T.; Wang, Y.; Slid, S.; Karelson, M. QSPR Treatment of Solvent Scales. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 684–691.

- (7) Katritzky, A. R.; Tamm, T.; Wang, Y.; Karelson, M. A Unified Treatment of Solvent Properties. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 692–698.
- (8) Abraham, M. H. Physicochemical and Biochemical Processes. *Chem. Soc. Rev.* **1993**, 22, 73–83.
- (9) Abraham, M. H.; Whiting, P. S.; Carr, P. W.; Ouyang, H. Hydrogen Bonding. Part 45. The Solubility of Gases and Vapours in Methanol at 298K: an LFER Analysis. *J. Chem. Soc., Perkin Trans. 2* **1998**, 1385–1390.
- (10) Abraham, M. H.; Whiting, G. S.; Shuely, W. J.; Doherty, R. M. The solubility of Gases and Vapours in Ethanol—the Connection Between Gaseous Solubility and Water–Solvent Partition. *Can. J. Chem.* **1998**, 76, 703–709.
- (11) Kurtz, H. A.; Stewart, J. J. P.; Dieter, K. M. Calculation of the nonlinear optical properties of molecules. *J. Comput. Chem.* **1990**, 11, 82–87.
- (12) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **1996**, 96, 1027–1043.
- (13) Roy, A. B.; Basak, S. C.; Harriss, D. K.; Magnuson, V. R. Neighborhood Complexities and Symmetry of Chemical Graphs and Their Biological Applications. In *Mathematical Modelling in Science and Technology*; Avula, X. J. R., Kalman, R. E., Liapis, A. I., Rodin, E. Y., Eds.; Pergamon Press: New York, 1984; pp 745–750.
- (14) Fowler, F. W.; Katritzky, A. R.; Rutherford, R. J. D. The Correlation of Solvent Effects on Physical and Chemical Properties. *J. Chem. Soc., Part B* **1971**, 460–469.
- (15) Koppel, I. A.; Palm, V. A. The Influence of the Solvent Organic Reactivity. In *Advances in Linear Free Energy Relationships*; Chapman, N. B., Shorter, J., Eds.; Plenum Press: London, New York, 1972; pp 203–247.
- (16) Kamlet, M. J.; Abboud, J.-L. M.; Abraham, M. H.; Taft, R. W. Linear Solvation Energy Relationships. 23. A Comprehensive Collection of the Solvatochromic Parameters, π^* , α , and β , and Some Methods for Simplifying the Generalized Solvatochromic Equation. *J. Org. Chem.* **1983**, 48, 2877–2887.
- (17) Krygowsky, T. M.; Fawcett, W. R. A Characteristic Vector Analysis of Solvent Effects for Thermodynamic Data. *Can. J. Chem.* **1976**, 54, 3283–3292.
- (18) Swain, C. G.; Swain, M. S.; Powell, A. I.; Alunni, S. Solvent Effects of Chemical Reactivity. Evaluation of Anion and Cation Solvation Components. *J. Am. Chem. Soc.* **1983**, 105, 502–513.
- (19) Mayer, U. NMR–Spectroscopic Studies on Solute–Solvent and Solute–Solute Interactions. In *Ions and Molecules in Solution*; Tanaka, N., Ohtaki, H., Tamamushi, R., Eds.; Elsevier: Amsterdam, 1983; p 219.
- (20) Dougherty, R. C. A Perturbation Molecular Orbital Approach to Ionic Solvation. A New Multiparameter Solvent Correlation Based on Solvent Ionization Potentials and Electron Affinities. *Tetrahedron Lett.* **1975**, 385–388.
- (21) PCMODEL, *Molecular modeling package*; Serena Software, Bloomington, 1992.
- (22) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, 107, 3902–3909.
- (23) Baker, J. An Algorithm for the Location of Transition States. *J. Comput. Chem.* **1986**, 7, 385–395.
- (24) Stewart, J. J. P. *MOPAC Program Package*; QCPE 1989, No. 455.
- (25) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, 96, 1027–1043.
- (26) Katritzky, A. R.; Lobanov, V.; Karelson, M. Physical Properties from Structure. *Chem. Soc. Rev.* **1995**, 279–287.
- (27) Katritzky, A. R.; Lobanov, V.; Karelson, M.; Murugan, R.; Grendze, M. P.; Toomey, J. E. Comprehensive Descriptors For Structural and Statistical Analysis. 1. Correlations Between Structure and Physical Properties of Substituted Pyridines. *Rev. Roum. Chim.* **1996**, 41, 851–867.
- (28) Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics. *J. Phys. Chem.* **1996**, 100, 10400–10407.
- (29) Draper, N. R.; Smith, H. *Applied Regression Analysis*; Wiley: New York, 1966.
- (30) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *CODESSA, User's Manual*; University of Florida & Tartu University: 1974.

CI000124V