

# A model based QSPR analysis of the unified non-specific solvent polarity scale



Lan Mu, Russell S. Drago\* and David E. Richardson

Department of Chemistry, University of Florida, Gainesville, FL 32611-7200, USA

A Model Based QSPR approach, MQSPR, has been employed to estimate the  $S'$  parameters of the Unified Solvent Polarity Scale. MQSPR selects descriptors, *a priori* to the correlation analysis, that have potential meaning in the context of the measured property. A successful two-parameter correlation ( $R^2 = 0.9587$ ) is developed from 48 diverse molecules using two orthogonal descriptors, the dipolar density (given by the total dipole moment of the molecule divided by the molecular volume) and the reciprocal of the HOMO–LUMO energy gap. The correlation equation allows confident estimations of  $S'$  from quantum mechanical calculations, assists in the interpretation of non-specific solvation and enables one to analyse solvent conformational-dependent solvation influences.

## Introduction

Early approaches to the understanding of solvent effects used simple macroscopic solvent properties, such as the static relative permittivity, permanent dipole moment and refractive index, or combinations of these functions.<sup>1</sup> Such approaches fail, especially when specific solute–solvent interactions exist, *e.g.* hydrogen bonding and charge transfer. The solvation process involves solvent reorganization to form cavities that accommodate the solute molecules with stabilization resulting from the interaction of the solute dipole (and induced dipole) with that of the solvent. The dominant contribution involves the first solvation shell of solvent molecules, *i.e.* the internal relative permittivity of the cavity.<sup>2</sup> Solvent reorganization and induced dipole moments tend to create an internal permittivity that differs from the bulk relative permittivity.<sup>2</sup> Not only is the internal permittivity of the organized solvent region difficult to quantify but so are the dimensions of the solvent cavity surrounding different solute molecules. Assumptions are invariably made to estimate these quantities.<sup>2</sup> It has not been possible to define solvent polarity in terms of macroscopic solvent properties, and as a result empirical solvent polarity scales have evolved as an alternative approach for predicting or analysing solvent effects.

Numerous empirical polarity scales have been reported.<sup>2–4</sup> A disadvantage of most of these scales<sup>3</sup> is that in addition to non-specific effects, they contain specific effects which are unique to the probe (solute) used to develop each scale. As a result, each scale differs for reasons that are not obvious, and this inhibits the understanding of solvent influence on new solutes.

In earlier reports, a unified solvation model was offered to treat both specific and non-specific effects [eqn. (1)]:<sup>2,4,5</sup>

$$\Delta\chi = PS' + E_A E_B + C_A C_B + W \quad (1)$$

In eqn. (1),  $\Delta\chi$  is the solvent-dependent physicochemical property,  $P$  is the susceptibility of the solute probe to polarity,  $S'$  is the solvent polarity derived from experimental observations,  $E_A$  and  $E_B$  are the electrostatic specific interaction parameters of the acid and base,  $C_A$  and  $C_B$  are the covalent parameters of the acid and base and  $W$  is a constant that is the value of  $\Delta\chi$  when  $E_A = C_A = P = 0$ . This model separates solvent effects into non-specific ( $PS'$ ) and specific ( $E_A E_B + C_A C_B$ ) interactions. The former arise from non-specific electrostatic forces involving bulk and internal solvent permittivity influences on charged ions, non-polar and dipolar solutes. The latter comprises electron-pair, donor–acceptor interactions including hydrogen-bonding and  $\pi$ – $\pi^*$  charge transfer. For less polar

solvents, the  $S'$  values correlate well with Hildebrand's solubility parameter.<sup>5</sup> For polar solvents, the  $S'$  values fit well in a correlation equation with Hansen's solubility parameters.<sup>5</sup> Thus,  $S'$  encompasses both solvent dispersion and dipolar effects.

The  $S'$  scale encompasses a wide variety of solutes and solvents, but excludes specific solute–solvent interactions, thus providing a scale of non-specific solvent polarity. Disparities between  $S'$  and other scales will result if specific interactions exist in the latter scales. With confidence in this model, deviations resulting from predictions using  $S'$ ,  $E_B$  and  $C_B$  are analysed to determine the unusual effects that cause the deviation, instead of searching for another scale to obtain correlations that happen to work for the set of solvents selected.<sup>2</sup> The same  $S'$  value can be used for a wide range of solute shapes and sizes.<sup>2</sup>

In addition to providing empirical solvation parameters, the  $S'$  parameters have been used with a cavity term to correlate enthalpies and free energies of solvation.<sup>5</sup> The trends in the magnitudes of the solute parameters lead to a dynamic cavity model for solvation in which the probe polarity is more important than the probe size in determining the cavity term.

It is not clear how the fundamental macroscopic properties of solvents influence their solvation abilities. Even if strong correlations could be found, experimental measurements of new solvent properties (relative permittivity, refractive index, spectral shifts, *etc.*) would have to be made to estimate  $S'$ . An approach which determines  $S'$  using fundamental molecular properties calculated by quantum mechanics would be valuable. A link between  $S'$  and calculated molecular microscopic properties would permit the estimation of  $S'$  for any molecule and the influence of these structural properties on solvation could be elucidated.

Famini has reported QSPR (quantitative structure–property relationships) treatments of solvent effects on physicochemical properties using theoretical descriptors<sup>6</sup> related to Kamlet–Taft solvation parameters. Recently a preliminary study of  $S'$  has been reported involving a search for the best multilinear regression for  $S'$  from a large number of descriptors programmed in CODESSA (Comprehensive Descriptors for Statistical and Structural Analysis).<sup>7</sup> A three-parameter model was obtained employing the following nonempirical descriptors: (i) the Average Structural Information Content (order 0); (ii) the Weighted Partial Negative Surface Area; (iii) the Hydrogen-bonding Acceptor Surface Area.<sup>8</sup> While this model functions as a predictive tool for  $S'$ , the intrinsic meaning of these parameters, as they relate to  $S'$ , is not readily appreciated.

**Table 1** The 29 diverse solvents and their  $S'$  values and the corresponding weights, refractive index ( $n$ ), relative permittivities ( $\epsilon$ ), and the two empirical functions<sup>a</sup>

Solvent	$S'$	$n$	$\epsilon$	$(1 - 1/n^2)/$ $MV \times 10^{-3}$	$1 - 1/\epsilon$	Wt. <sup>b</sup>
1,2-Dichlorobenzene	2.10	1.5510	9.93	5.19	0.8993	0.6
1,4-Dioxane chair	1.93	1.4220	2.21	5.93	0.5475	1
Butan-2-one	2.50	1.3790	18.50	5.30	0.9459	0.8
Acetone	2.58	1.3590	20.70	6.24	0.9517	1
Acetonitrile	3.00	1.3440	37.50	8.54	0.9733	1
Anisole	2.04	1.5180	4.33	5.17	0.7691	0.6
Benzene	1.73	1.5011	2.27	6.26	0.5595	1
Benzonitrile	2.63	1.5289	25.20	5.61	0.9603	1
Bromobenzene	2.10	1.5570	5.40	5.69	0.8148	0.4
Carbon disulfide	1.51	1.6320	2.64	10.40	0.6212	0.2
Carbon tetrachloride	1.49	1.4660	2.24	5.54	0.5532	1
Chlorobenzene	1.98	1.5240	5.62	5.60	0.8221	0.8
Cyclohexane	1.11	1.4262	2.02	4.71	0.5050	1
Di- <i>n</i> -butyl ether	1.58	1.3992	3.08	2.90	0.6753	0.6
Dichloromethane	2.08	1.3300	9.08	6.83	0.8899	0
Diethyl ether	1.73	1.3520	4.34	4.36	0.7693	1
Dimethyl sulfoxide	3.00	1.4780	46.48	7.64	0.9785	1
Ethyl acetate	2.15	1.3720	6.02	4.79	0.8339	1
<i>N,N</i> -Dimethylacetamide	2.70	1.4351	37.80	5.53	0.9735	1
<i>N,N</i> -Dimethylformamide	2.80	1.4269	36.71	6.57	0.9728	1
<i>n</i> -Decane	0.90	1.4121	1.99	2.56	0.4977	0.2
<i>n</i> -Hexane	0.68	1.3750	1.89	3.6	0.4709	0.2
Nitrobenzene	2.61	1.5562	34.82	5.74	0.9713	1
Nitromethane	3.07	1.3935	35.87	9.02	0.9721	0.8
Pyridine	2.44	1.5100	12.40	6.97	0.9194	1
Tetrahydrofuran	2.08	1.4070	7.58	6.10	0.8681	1
Toluene	1.66	1.4970	2.38	5.21	0.5797	0.8
Trichloromethane	1.74	1.4440	4.81	6.50	0.7919	0
Triethylamine	1.43	1.4010	2.42	3.52	0.5868	1

<sup>a</sup> MV is the molar volume of solvents. A weight of zero indicates these solvents invariably have contributions from specific interactions. <sup>b</sup> The weights are cited from ref. (5).

In this report, other non-empirical descriptors are selected *a priori* to have potential meaning in the context of solvation. Correlations using these selected properties will be referred to as MQSPR, a Model-Based QSPR Approach. Since application of a model-based correlation to the solvation parameters employs intrinsic molecular quantities related to the property, qualitative and quantitative predictions of the parameters can be made to meet experimental demands with more confidence.

## Methodology

Experimental  $S'$  values for 67 solvents were assembled from previous publications.<sup>5,9,10</sup> This set included saturated and unsaturated hydrocarbons, halogenated solvents, solvents containing halogen, cyano, nitro, amide, sulfide, mercapto, sulfone, phosphate, ester, ether and carbonyl groups as well as furan, pyran, dioxane, pyridine, aniline, quinoline, imidazole, pyrrolidinone and pyrazine rings. Structures of the solvents were drawn and pre-optimized by the MMX molecular mechanics method using the PCMODEL<sup>11</sup> program. The final geometry optimization of these compounds was performed on an IBM RISC/6000 model 320 computer using the semi-empirical quantum-chemical AM1 parametrization with the MOPAC 6.0 program which was modified by incorporation of a Kirkwood–Onsager self-consistent reaction field (SCRf).<sup>12</sup> A general relative permittivity (in our case that of water) was used in the reaction field tensor as in classical Kirkwood–Onsager applications. Since this is a scaled correction, similar results with different correlation coefficients are anticipated if gas phase structures were used. The MOPAC results for individual compound were loaded into the CODESSA program along with the  $S'$  data. Weighted fits are employed so that greater significance could be given to the more reliable  $S'$  parameters. Also, in order to set the calculated  $S'$  of DMSO at 3.0, the weight for this  $S'$  was set at 12.

A large number of descriptors (*ca.* 100) are calculated with

the CODESSA program. In QSPR, statistical procedures implemented in CODESSA select the relevant descriptors. The MQSPR approach makes an *a priori* determination of descriptors with physical significance in the context of the problem being addressed and selects the most meaningful combination that results from statistical analysis. For non-specific solvation, the only descriptors considered relevant involve molecular volume, surface area, polarizability, dipole moment and HOMO LUMO energy. These were calculated. Constitutional descriptors<sup>13</sup> such as numbers of atoms, bonds *etc.*, and descriptors that are solely associated with a specific constituent, such as the minimum atomic state energy for a certain atom, and the average valence of a certain atom *etc.* were not included for there is no physical basis for their relevance to solvation.

## Results and discussion

### Correlation to macroscopic properties

It was found that  $S'$  for 29 solvents (Table 1), for which  $\epsilon$  (the solvent relative permittivity) and  $\eta$  (the solvent refractive index) are available, could be fitted to the sum of two experimental functions,  $(1 - 1/\epsilon)$  and  $(1 - 1/\eta^2)/MV$ , where MV is the molar volume of the solvent. The weighted correlation gave eqn. (2)

$$S' = -(0.83 \pm 0.09) + (284 \pm 1)(1 - 1/\epsilon) + (126 \pm 16)(1 - 1/\eta^2)/MV \quad (2)$$

with an  $R^2$  of 0.978, an average absolute error of 0.09 and an  $F$ -value of 564. Since more  $\epsilon$  and  $\eta$  values are available than  $S'$  values, the correlation provides a useful equation for the estimation of approximate  $S'$  values. In addition, the empirical fit to these two functions indicate that the dipolarities, approximated by relative permittivities, and dispersion effects, approximated by the refractive index function, are important solvent properties contained in  $S'$ . The term  $(1 - 1/\epsilon)$  was used because this

**Table 2** The one-parameter correlations for the empirical set of 29 diverse solvents

Descriptor	$R^2$
Total dipole moment/molar volume	0.9059
Total dipole moment	0.8362
Image of the Onsager–Kirkwood solvation energy	0.8037
Molar volume	0.4359
$\alpha$ -Polarizability	0.2846
Molecular weight	0.2396
HOMO–LUMO energy gap	0.2132
LUMO energy	0.1592

function is the solvent dependent term of the Born solvation energy of an ion. The term  $(1 - 1/\eta^2)$  was used because it is proportional to that part of the Born solvation energy attributable to the purely electronic polarization of the solvent. Reciprocals of the refractive indices and relative permittivities have been utilized in an empirical modelling of Catalan's solvent polarity scale.<sup>14</sup>

### MQSPR

**Empirical solvent set.** The  $S'$  values for the 29 solvents used in the correlation with  $1 - 1/\epsilon$  and  $(1 - 1/\eta^2)MV$ , were selected as the primary working set for MQSPR. Table 1 provides the names of the solvents, the corresponding  $S'$  values and their weights. The refractive indices and relative permittivities of the solvents,<sup>15</sup> are also listed in Table 1 along with the two terms in eqn. (2). The weights, which indicate the reliability of the  $S'$  values, range from 0–1.0, and were loaded into CODESSA.

Table 2 gives the correlations found for  $S'$  to several descriptors. Various dipole moment functions to account for dipolarity of the solvents and other quantum-chemical descriptors to represent the polarizability of the solvents were attempted. A fair one-parameter correlation of  $S'$  to the total dipole moment of the molecule,  $\mu$ , is found indicating the dominant contribution of polarity in  $S'$ . The Onsager–Kirkwood solvation energy (denoted as  $E_{\text{Onsager}}$  in Table 2) is given by eqn. (3), where  $\epsilon$  is the

$$E_{\text{Onsager}} = -\frac{(\epsilon - 1)\mu^2}{(2\epsilon + 1)a^3} \quad (3)$$

macroscopic relative permittivity of the solvent and  $\mu$  is the total dipole moment of the solvent and  $a$  is the radius of the solvent cavity. This function does not fit as well as  $\mu$ . Descriptors that are associated with polarizability (Table 2), such as molecular volume, molecular weight,  $\alpha$  polarizability, HOMO–LUMO energy gap and LUMO energy give poor one-parameter correlations.

Prompted by the involvement of  $MV$  in the  $(1 - 1/\eta^2)/MV$  term of eqn. (2), a new descriptor, the magnitude of total dipole moment of a molecule ( $\mu$ ) divided by the calculated molecular volume ( $MV_{\text{cal}}$ ), was constructed. The resulting composite descriptor,  $\mu/MV_{\text{cal}}$ , substantially improved the one-parameter dipole moment correlation (from  $R^2 = 0.8362$  to  $R^2 = 0.9059$ ). The descriptor  $\mu/MV_{\text{cal}}$ , referred to as the dipole density, is conceptually similar to the charge density of ionic species. The dipole density function is physically justified since the influence of a molecular dipole moment of a given magnitude on solvation will vary with molecular size, *i.e.* more solvent molecules can reside in the first solvation shell of the solute as the solvent molecular volume decreases. Thus, a greater number of smaller solvent molecules solvating a given polar solute will lead to a larger total electrostatic interaction of the solvent with the solute than a fewer number of larger solvent molecules with the same dipole moment.

A number of descriptors, known to be associated with polarizability, were attempted along with  $\mu/MV_{\text{cal}}$  in two-parameter correlations to  $S'$ . When any one of seven descriptors were added to  $\mu/MV_{\text{cal}}$ , an improved correlation resulted. Table 3

**Table 3** The two-parameter correlations of  $S'$  for the 29 diverse solvents<sup>a</sup>

Descriptors	$R^2$	$F$	$a_0$	$a_1$	$a_2$
$\mu/MV$	0.8957	232	1.54	24.02	
$\mu/MV + E_{\text{HOMO-LUMO}}$	0.9349	175	2.55	22.80	−0.09
$\mu/MV + 1/E_{\text{HOMO-LUMO}}$	0.9303	174	0.56	25.58	11.11
$\mu/MV + \alpha/MV$	0.9144	139	0.81	24.82	1.27
$\mu/MV + E_{\text{LUMO}}$	0.9067	126	1.64	21.78	−0.09
$\mu/MV + E_{\text{HOMO-LUMO}}/MV$	0.9027	121	1.67	25.10	−1.46
$\mu/MV + {}^0\text{AIC}$	0.9002	117	1.37	16.01	0.63
$\mu/MV + E_{\text{HOMO}}$	0.8970	106	2.46	24.36	0.10
$\mu/MV + \alpha$	0.8957	112	1.49	24.15	0.00 <sub>1</sub>

<sup>a</sup>  $a_0$  is the intercept,  $a_1$  is the coefficient of  $\mu/MV$  and  $a_2$  is the coefficient of the second descriptor.  $\alpha$  is the polarizability.

lists eight second descriptors, the corresponding two-parameter correlation coefficients, the t-test, the  $F$ -values and the coefficients ( $a_1$  and  $a_2$ ) of the two parameters. The best two-parameter correlations for the 29 solvents involve the molecular volume weighted total dipole of the molecule ( $\mu/MV_{\text{cal}}$ ) and functions of the HOMO–LUMO energy gap (either  $E_{\text{HOMO-LUMO}}$  or  $1/E_{\text{HOMO-LUMO}}$ ). A larger HOMO–LUMO energy gap corresponds to a smaller polarizability and a negative coefficient results for the fit to  $E_{\text{HOMO-LUMO}}$ . The reciprocal of the HOMO–LUMO energy gap directly relates to polarizability and is the more reasonable descriptor for the dispersion energy of solvents.<sup>16</sup> Eqn. (4) results with  $R^2$  of 0.9303

$$S' = (0.56 \pm 0.27) + (25.58 \pm 1.37)\mu/MV + (11.11 \pm 3.09)1/E_{\text{HOMO-LUMO}} \quad (4)$$

an  $F$  of 174 and an average absolute error of 0.16. Among the solvents for which the  $S'$  values are weighted as 1 on the 0–1 scale, 1,4-dioxane is grossly under-predicted and will be discussed later. Thus, the two principal intrinsic components of  $S'$ , dipolarity and polarizability, are represented by the dipole density of the molecule and the reciprocal of the HOMO–LUMO energy gap. The intercorrelation between the two descriptors gives an  $R^2$  of 0.005 so each descriptor describes a single and orthogonal molecular characteristic. These descriptors are consistent with the literature<sup>1–3,6</sup> recognition of the importance of polarizability and dipolarity contributions to solvation. Since these two parameters can be easily obtained through quantum-chemical calculations, the working set was next expanded to include more solvents.

**Evaluation of MOPAC calculated properties.** Ideally, all 67 solvents with known  $S'$  values (Table 4) could be included in the working set. However, caution has to be taken, since in previous reports<sup>17</sup> the AM1 parametrization yielded an inadequate geometry and partial charge distribution for nitrogen, sulfur and phosphorus containing molecules. The quality of the calculated water reaction field, total dipole moments and molecular volumes calculated with MOPAC are investigated by comparing them to the experimental dipole moments and molar volumes<sup>15a,18</sup> providing the results in Table 4. The correlation between the 66 known experimental and calculated dipole moments gave  $R^2 = 0.9240$ ,  $F = 778$ , and an average absolute error of 0.28 D. Solvents such as 4-butyrolactone, *N*-methylimidazole and those containing nitro groups are recognized as the most distinct outliers. Sulfur containing solvents are accurately predicted. Nitrobenzene has the largest absolute error of 2.25 D. All predicted values that exceed 2.5 times the average absolute error were omitted from the final correlation analysis. Thus, five solvents, labelled as *a* in Table 4 were removed from the working set. The correlation between experimental and calculated dipole moments for the remaining set of 61 solvents led to eqn. (5) with an  $R^2 = 0.9606$ ,  $F = 143$  and an absolute error of 0.20 D (Fig. 1).

$$\mu_{\text{exp}} = (0.054 \pm 0.066) + (0.97 \pm 0.03)\mu_{\text{calc}} \quad (5)$$

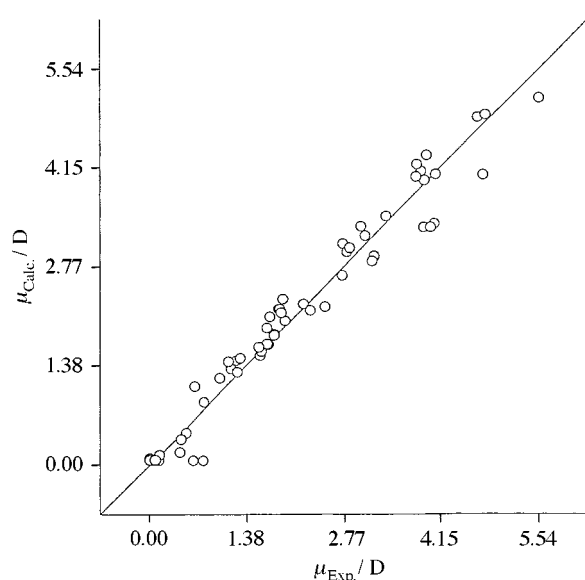
**Table 4** The experimental (exp) and calculated (calc.) dipole moments and molar volumes and the corresponding absolute errors ( $\Delta s$ )

Solvent	Dipole moment			Molar volume		
	Exp.	Calc.	$\Delta s$	Exp.	Calc.	$\Delta s$
1,1,1-Trichloroethane	1.78	1.81	0.03	100.3	86.3	14.0
<i>cis</i> -1,1,2-Trichloroethane	n/a	2.57	n/a	93.0	86.1	6.9
<i>trans</i> -1,1,2-Trichloroethane <sup>b</sup>	1.56	1.26	0.30	93.0	86.7	6.3
1,1,2-Trichloroethene	0.77	0.84	0.07	90.1	79.7	10.4
1,2-Dichlorobenzene	2.50	2.21	0.29	113.2	112.8	0.4
boat 1,4-Dioxane	n/a	1.79	n/a	85.7	85.7	0.0
chair 1,4-Dioxane <sup>b</sup>	0.42	0.12	0.30	85.7	85.9	0.2
Butan-2-one	2.76	3.12	0.36	90.2	81.1	9.1
3-Methylsulfolane	4.80	4.99	0.19	113.0	117.7	4.7
Pentan-3-one	2.82	3.00	0.18	106.4	97.8	8.6
4-Butyrolactone <sup>a</sup>	4.18	5.04	0.86	76.5	79.6	3.1
4-Methylpyridine	2.75	2.67	0.08	98.1	100.2	2.1
Acetone	2.86	3.06	0.20	73.9	64.2	9.7
Acetonitrile	3.92	3.35	0.57	52.7	47.7	5.0
Acetophenone	3.02	3.37	0.35	117.4	123.4	6.0
Anisole	1.25	1.44	0.19	109.2	113.3	4.1
Benzene	0.03	0.00	0.03	89.5	87.1	2.4
Benzonitrile	3.93	4.03	0.10	103.0	105.7	2.7
Bromobenzene	1.70	1.66	0.04	105.6	105.5	0.1
Butyl acetate <sup>c</sup>	1.84	2.18	0.34	132.6	123.7	8.9
Carbon disulfide	0.12	0.00	0.12	60.6	67.8	7.2
Carbon tetrachloride	0.00	0.01	0.01	97.1	83.6	13.5
Chlorobenzene	1.58	1.51	0.07	102.3	100.2	2.1
Cyclohexane	0.00	0.00	0.00	108.9	101.1	7.8
Cyclohexanone	3.08	3.23	0.15	104.1	104.0	0.1
Di-isopropyl ether <sup>a</sup>	1.13	1.42	0.29	141.8	121.3	20.5
Di- <i>n</i> -butyl ether	1.17	1.31	0.14	170.3	154.8	15.5
Dichloromethane	1.60	1.56	0.04	64.4	55.9	8.5
Diethyl ether	1.30	1.47	0.17	104.7	87.9	16.8
Dimethylaniline	1.68	1.67	0.01	127.6	133.3	5.7
Dimethyl sulfoxide	3.96	4.40	0.44	71.3	73.8	2.5
Ethyl acetate <sup>c</sup>	1.78	1.80	0.02	98.5	90.0	8.5
Ethyl formate	1.93	2.00	0.07	80.8	72.4	8.4
Hexamethylphosphoramide	5.54	5.47	0.07	175.7	177.2	1.5
Hexyl acetate <sup>c</sup>	1.86	2.18	0.32	n/a	157.2	n/a
Methyl acetate <sup>c</sup>	1.68	1.90	0.22	79.9	73.2	6.7
<i>N,N</i> -Diethylformamide	3.88	4.16	0.28	n/a	111.7	n/a
<i>N,N</i> -Dimethylacetamide	3.81	4.09	0.28	93.0	93.1	0.1
<i>N,N</i> -Dimethylcyanamide	4.77	4.11	0.66	n/a	77.3	n/a
<i>N,N</i> -Dimethylformamide	3.82	4.27	0.45	77.4	77.7	0.3
<i>N</i> -Methylimidazole <sup>a</sup>	3.71	4.53	0.82	n/a	85.4	n/a
<i>N</i> -Methylpyrrolidinone	4.09	4.12	0.03	96.7	100.2	3.5
<i>n</i> -Butyronitrile	4.07	3.41	0.66	87.9	81.7	6.2
<i>n</i> -Decane <sup>b</sup>	0.00	0.03	0.03	195.3	180.1	15.2
<i>n</i> -Heptane <sup>b</sup>	0.00	0.01	0.01	147.0	129.4	17.6
<i>n</i> -Hexane <sup>a</sup>	0.00	0.00	0.00	131.3	113.0	18.3
<i>n</i> -Nonane <sup>a</sup>	0.00	0.01	0.01	181.4	163.0	18.4
<i>n</i> -Pentane <sup>a</sup>	0.07	0.01	0.06	119.5	95.0	24.5
Nitrobenzene <sup>a</sup>	3.93	6.18	2.25	102.7	114.6	11.9
Nitroethane <sup>a</sup>	3.65	4.85	1.20	72.0	72.9	0.9
Nitromethane <sup>a</sup>	3.46	4.56	1.10	54.9	56.2	1.3
Propionitrile	4.02	3.36	0.66	70.9	64.7	6.2
Propyl acetate <sup>c</sup>	1.88	2.12	0.24	115.7	106.4	9.3
Propylene carbonate	n/a	6.61	n/a	n/a	87.6	n/a
Pyridine	2.19	2.25	0.06	80.8	82.8	2.0
Quinoline	2.29	2.16	0.13	118.5	128.5	10.0
Tetrahydrofuran	1.72	2.06	0.34	81.9	77.4	4.5
Tetrahydropyran	1.56	1.63	0.07	n/a	93.5	n/a
Tetrahydrothiophene	1.90	2.33	0.43	88.4	86.4	2.0
Tetramethylene sulfone	4.69	4.95	0.26	95.3	101.1	5.8
Tetramethylurea	3.38	3.52	0.14	n/a	122.5	n/a
Thiophene	0.51	0.40	0.11	79.5	81.8	2.3
Toluene	0.43	0.31	0.12	106.6	104.3	2.3
Tributyl phosphate <sup>b</sup>	3.21	2.93	0.28	n/a	240.0	n/a
Trichloromethane	1.01	1.18	0.17	80.7	69.7	11.0
Triethylamine	0.63	1.06	0.43	139.7	125.3	14.4
Triethyl phosphate <sup>b</sup>	3.21	2.94	0.27	n/a	167.7	n/a
Trimethylbenzene	0.13	0.08	0.05	139.5	137.0	2.5
Trimethyl phosphate <sup>b</sup>	3.18	2.87	0.31	n/a	118.1	n/a

<sup>a</sup> Solvents that are removed from the working set due to large deviations between the experimental and calculated dipole moments and molar volumes. <sup>b</sup> Solvents for which the predicted  $S'$  values exceed 2.5 times the average absolute error in the primary  $S'$  fitting, and are removed from the working set. <sup>c</sup> Solvents with large conformation-dependent total dipole moment differences but which may not be able to rearrange in the solute vicinity, and are not included in the final working set.

**Table 5** The calculated dipole moments ( $\mu$ ), the  $S'$  values and the energies relative to global minimum ( $\Delta E$ ) for different conformations of solvents

Solvent	$\mu/D$	$S'_{\text{Calc}}$	$\Delta E/\text{kcal mol}^{-1}$	$S'_{\text{Exp}}$	Weight
boat 1,4-Dioxane	1.79	1.95	2.2		
chair 1,4-Dioxane	0.12	1.48	0.0	1.93	1.0
<i>cis</i> -1,1,2-Trichloroethane	2.57	1.87	0.3		
<i>trans</i> -1,1,2-Trichloroethane	1.26	2.22	0.0	2.35	0.2
boat Cyclohexane	0.005	1.38	3.2		
chair Cyclohexane	0.002	1.37	0.0	1.11	1.0
boat Cyclohexanone	3.25	2.30	0.1		
chair Cyclohexanone	3.23	2.30	0.0	2.35	0.6
Butyl acetate	5.01	2.46	3.5		
Butyl acetate'	2.18	1.90	0.0	1.99	0.4
Di-isopropyl ether	1.46	1.72	0.4		
Di-isopropyl ether'	1.42	1.71	0.0	1.76	0.8
Di- <i>n</i> -butyl ether	1.43	1.65	1.7		
Di- <i>n</i> -butyl ether'	1.31	1.63	0.0	1.58	0.6
Diethyl ether	1.65	1.86	0.4		
Diethyl ether'	1.47	1.82	0.0	1.73	1.0
Ethyl acetate	5.03	2.82	2.0		
Ethyl acetate'	1.80	1.95	0.0	2.15	1.0
Hexamethylphosphoramide	5.23	2.30	1.6		
Hexamethylphosphoramide'	5.04	2.27	0.0	2.52	0.6
Hexyl acetate	5.47	2.26	2.8		
Hexyl acetate'	2.18	1.82	0.0	1.94	0.4
Methyl acetate	5.00	3.12	1.3		
Methyl acetate'	1.90	2.09	0.0	2.35	0.4
Nitroethane	4.89	6.35	0.1		
Nitroethane'	4.85	3.10	0.0	2.78	0.2
Propyl acetate	5.04	2.62	2.3		
Propyl acetate'	2.12	1.96	0.0	2.05	0.4
boat Propylene carbonate	6.62	3.25	0.0		
chair Propylene carbonate	6.61	3.24	0.0	3.10	0.6
Tetrahydrofuran	2.08	2.05	0.1		
Tetrahydrofuran'	2.08	2.05	0.0	2.08	1
boat Tetrahydropyran	1.68	1.85	2.8		
chair Tetrahydropyran	1.63	1.82	0.0	1.98	0.4
Triethylamine	1.13	1.72	0.5		
Triethylamine'	1.06	1.70	0.0	1.43	1
Tributyl phosphate	5.38	2.00	169.1		
Tributyl phosphate'	2.93	1.77	0.0	2.30	0.4
Triethyl phosphate	5.35	2.38	169.4		
Triethyl phosphate'	2.94	1.91	0.0	2.55	0.6
Trimethyl phosphate	5.64	2.95	151.7		
Trimethyl phosphate'	2.87	2.07	0.0	2.79	0.2

**Fig. 1** The calculated vs. experimental dipole moment of 61 diverse solvents

Propylene carbonate was not included in the fit to eqn. (5) because experimental data is not available. Nevertheless, this solvent was kept in the working set to correlate to  $S'$  since it consists of atoms that the AM1 parameterization can handle.

The correlation between the 57 known<sup>15a</sup> experimental molar volumes,  $MV_{\text{exp}}$ , and the calculated molecular volumes  $MV_{\text{cal}}$  (in units of  $\text{ml mol}^{-1}$ ) has an  $R^2 = 0.9277$ ,  $F = 705$ , and an average absolute error of  $7.26 \text{ ml mol}^{-1}$ . *n*-Hexane, *n*-nonane, *n*-pentane and diisopropyl ether are outliers which have absolute errors that exceeded 2.5 times the average absolute error ( $18.1 \text{ ml mol}^{-1}$ ). Removing these solvents (also labelled as *a* in Table 4) from the set, led to eqn. (6) with an improved  $R^2 =$

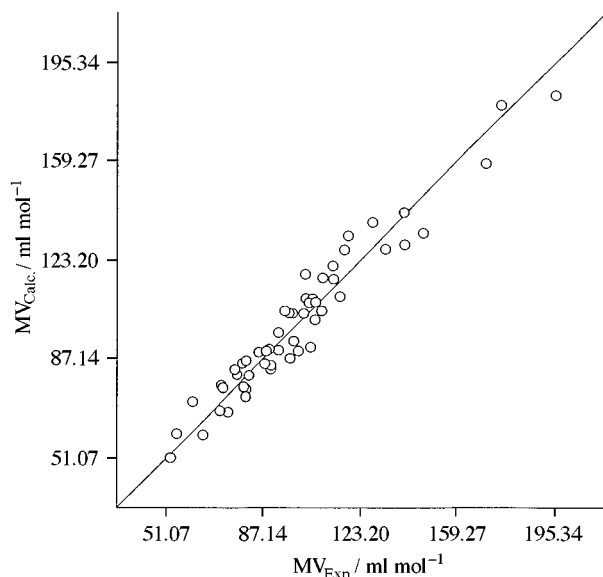
$$MV_{\text{exp}} = (3.40 \pm 3.65) + (1.00 \pm 0.04)MV_{\text{cal}} \quad (6)$$

$0.9372$ ,  $F = 761$ , and an average absolute error of  $6.26 \text{ ml mol}^{-1}$  (Fig. 2).

**Extension of the model to additional solvents.** The good correlations to  $MV$  and  $\mu$  encouraged a data fit of the enlarged set of  $S'$  values to the calculated dipole densities and the reciprocal of HOMO–LUMO energy gap. Nine solvents (labelled as *a* in Table 4) are removed from the working data set as a result of the deviation found in the experimental dipole moment and molar volume correlations. The remaining 58 solvents were submitted for multilinear regression analysis. The  $R^2$  that resulted was  $0.9018$  with  $F = 252$  and an average absolute error of  $0.17$ . Solvents whose error exceeds 2.5 times the average absolute error include 1,4-dioxane, 1,1,2-trichloroethane, *n*-decane, *n*-heptane, tributyl phosphate, triethyl phosphate and trimethyl phosphate. Alkanes do not have very accurate  $S'$  values because of the aggregation of most spectral probes in these solvents. The three phosphates, whose  $S'$  values are not

**Table 6** The two-parameter correlations for  $S'$  of the working set (48 solvents) and the comparing set (48 solvents)

Set	Descriptor	$X \pm \Delta X$	t-test	$R^2$	$F$
Working set	Intercept	$0.60 \pm 0.15$	4.01		
	$\mu/\text{MV}$ (calc.)	$24.08 \pm 0.77$	31.32	0.9201	529
	$1/E_{\text{HOMO-LUMO}}$	$10.46 \pm 1.61$	6.50	0.9587	523
Comparing set	Intercept	$1.28 \pm 0.14$	9.16		
	$\mu/\text{MV}$ (exp.)	$14.98 \pm 0.76$	32.98	0.9572	1051
	$1/E_{\text{HOMO-LUMO}}$	$3.49 \pm 1.53$	2.28	0.9615	574

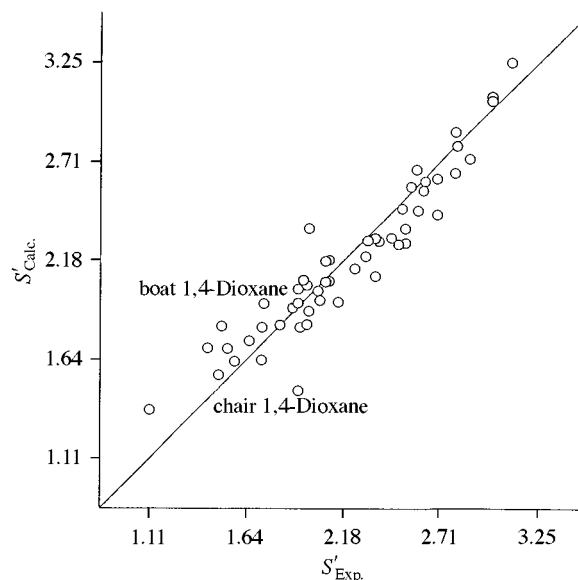
**Fig. 2** The calculated vs. experimental molar volume of 53 diverse solvents

well known and also have an error in their calculated dipole moments ranging from 0.27–0.31 D (Table 4), are assigned weights between 0.2–0.6. Both 1,1,2-trichloroethane and 1,4-dioxane have large conformation-dependent total dipole moment differences, and these solvents will be discussed in detail later. The regression analysis for the remaining 51 solvents gave eqn. (7) with an  $R^2 = 0.9536$ ,  $F = 493$  and an average absolute error of 0.11.

$$S' = (0.73 \pm 0.15) + (23.68 \pm 0.77)\mu/\text{MV} + (9.26 \pm 1.65)1/E_{\text{HOMO-LUMO}} \quad (7)$$

This equation compares favourably with the correlation shown in Table 3 for the set of 29 solvents that were first analysed.

**Conformational effects.** The predicted  $S'$  value for 1,4-dioxane, using eqn. (7), yielded an absolute error of 0.45. This is unacceptable since the observed  $S'$  for 1,4-dioxane is well known. Correlations with empirical solvent polarity scales<sup>14</sup> often encounter problems with 1,4-dioxane. The dipole moment of 1,4-dioxane is conformation dependent. The replacement of a solvent molecule in a solvent cluster by a solute molecule changes the polarity experienced by the solvent cluster in the region surrounding the solute. This dipolar interaction could induce a conformational change in the solvent. Even though our AM1 calculation incorporates a reaction field, this correction is merely an enhancement of the partial charge and dipole moment of all molecules,<sup>17</sup> and does not include changes in the dipole moment of a solvent due to changes in molecular conformation induced by solvent–solute interactions. Such changes can be simulated by varying the molecular conformation, which is easily accomplished with quantum chemical calculations. The boat conformation of 1,4-dioxane was calculated leading

**Fig. 3** The calculated vs. experimental  $S'$  of 53 diverse solvents (including chair and boat 1,4-dioxane) using the two-parameter correlation eqn. (7)

to a dipole moment of 1.8 D, a  $\mu/\text{MV}$  of 0.0209 and an  $E_{\text{HOMO-LUMO}}$  of 12.94. Substituting these properties for dioxane in its boat conformation into eqn. (7) leads to an  $S'$  value that falls on the regression line (Fig. 3).

The boat conformation of 1,4-dioxane is not the predominant species in solution and MOPAC calculations show that the chair conformation is 2.1 kcal mol<sup>-1</sup> more stable than the boat. However, in the vicinity of a solute molecule, the added dipole–dipole interaction of the solvent with the solute can provide the energy to change the conformation of the dioxane molecules that form the cavity wall. Alternatively, the orientation of dioxane in the cavity could have one of the oxygens pointing towards the solute and the other away so the dipole–dipole solute–solvent interaction is better approximated by the boat than the zero total dipole moment of the chair form. For solutes that are not very polar, the presence of dioxane molecules with both the chair and boat orientation in the cavity walls could lead to varying values of  $S'$  causing deviations that are less than that predicted by  $S'$  for the boat conformation. The dioxane system illustrates an important advantage of calculations for exploring peculiar solvent effects that cannot be studied with experiments. Conformation dependent variation of solvent molecules is one reason that the macroscopic neat solvent functions and parameters proposed by Kirkwood and Onsager fail.

MOPAC calculations showed that both *trans* and *cis* conformations of 1,1,2-trichloroethane have similar energies. Following the same argument made for dioxane, the bulk macroscopic properties of the *trans* derivative do not represent the solvation behaviour because the polar solute–solvent interactions are maximized with the *cis* conformation. This explanation is supported by the fact that the *trans* conformation fits the experimental dipole moment but not  $S'$  (Table 5) while

**Table 7** A comparison of the experimental and predicted  $S'$  of 67 compounds

No.	Solvent	Formula	$S'_{\text{exp}}$	$S'_{\text{calc}}$	$\mu/\text{MV}$	$E_{\text{HOMO-LUMO}}$	Wt. <sup>d</sup>
1	1,1,1-Trichloroethane	$\text{Cl}_3\text{C-CH}_3$	1.93 <sup>a</sup>	2.00	0.0210	11.72	0.2
2	<i>cis</i> -1,1,2-Trichloroethane	$\text{ClH}_2\text{CHCl}_2$	2.35 <sup>c</sup>	2.20	0.0298	11.83	0.2
2'	<i>trans</i> -1,1,2-Trichloroethane	$\text{ClH}_2\text{CHCl}_2$	2.35 <sup>c</sup>	1.84	0.0146	11.75	0.2
3	1,1,2-Trichloroethene	$\text{Cl}_2\text{C=CHCl}$	1.90 <sup>a</sup>	1.91	0.0105	9.90	0.4
4	1,2-Dichlorobenzene	$1,2\text{-Cl}_2\text{C}_6\text{H}_4$	2.10	2.17	0.0196	9.48	0.6
5	boat 1,4-Dioxane	$\text{O}(\text{CH}_2\text{CH}_2)_2\text{O}$	1.93 <sup>a</sup>	1.43	0.0013	13.03	1
5'	chair 1,4-Dioxane	$\text{O}(\text{CH}_2\text{CH}_2)_2\text{O}$	1.93 <sup>a</sup>	1.91	0.0209	12.94	1
6	Butan-2-one	$\text{CH}_3\text{C}(\text{O})\text{C}_2\text{H}_5$	2.50 <sup>a</sup>	2.44	0.0385	11.41	0.8
7	3-Methylsulfolane	$\text{C}_5\text{H}_{10}\text{SO}_2$	2.55	2.57	0.0424	11.04	0.4
8	Pentan-3-one	$\text{C}_5\text{H}_9\text{COC}_2\text{H}_5$	2.37	2.26	0.0307	11.32	0.5
9	4-Butyrolactone	$\text{C}_4\text{H}_6\text{O}_2$	2.86 <sup>a</sup>	2.97	0.0634	12.33	1
10	4-Methylpyridine	$4\text{-CH}_3\text{C}_5\text{H}_4\text{N}$	2.31	2.27	0.0266	10.10	0.8
11	Acetone	$(\text{CH}_3)_2\text{CO}$	2.58	2.66	0.0476	11.48	1
12	Acetonitrile	$\text{CH}_3\text{CN}$	3.00	3.04	0.0703	14.01	1
13	Acetophenone	$\text{C}_6\text{H}_5\text{C}(\text{O})\text{CH}_3$	2.52	2.35	0.0273	9.58	0.8
14	Anisole	$\text{C}_6\text{H}_5\text{OCH}_3$	2.04	2.01	0.0127	9.49	0.6
15	Benzene	$\text{C}_6\text{H}_6$	1.73	1.62	0.0000	10.20	1
16	Benzonitrile	$\text{C}_6\text{H}_5\text{CN}$	2.63	2.61	0.0382	9.57	1
17	Bromobenzene	$\text{C}_6\text{H}_5\text{Br}$	2.10	2.06	0.0158	9.68	0.4
18	Butyl acetate	$\text{CH}_3\text{C}(\text{O})\text{OBu}$	1.99	1.87	0.0176	12.29	0.4
19	Carbon disulfide	$\text{CS}_2$	1.51	1.83	0.0000	8.52	0.2
20	Carbon tetrachloride	$\text{CCl}_4$	1.49	1.53	0.0001	11.26	1
21	Chlorobenzene	$\text{C}_6\text{H}_5\text{Cl}$	1.98	2.04	0.0150	9.73	0.8
22	Cyclohexane	$\text{C}_6\text{H}_{12}$	1.11	1.33	0.0000	14.58	1
23	Cyclohexanone	$(\text{CH}_2)_5\text{CO}$	2.35 <sup>a</sup>	2.29	0.0314	11.22	0.6
24	Di-isopropyl ether	$\text{Pr}^i_2\text{O}$	1.76	1.68	0.0120	13.25	0.8
25	Di- <i>n</i> -butyl ether	$(n\text{-C}_4\text{H}_9)_2\text{O}$	1.58	1.61	0.0092	13.23	0.6
26	Dichloromethane	$\text{CCl}_2\text{H}_2$	2.08	2.15	0.0280	11.98	0
27	Diethyl ether	$(\text{C}_2\text{H}_5)_2\text{O}$	1.73	1.82	0.0187	13.57	1
28	Dimethylaniline	$\text{C}_6\text{H}_5\text{N}(\text{CH}_3)_2$	1.96 <sup>a</sup>	2.07	0.0125	8.92	0.2
29	Dimethyl sulfoxide	$(\text{CH}_3)_2\text{SO}$	3.00	3.04	0.0596	10.38	1
30	Ethyl acetate	$\text{CH}_3\text{C}(\text{O})\text{OC}_2\text{H}_5$	2.15	1.92	0.0199	12.40	1
31	Ethylformamide	$\text{HC}(\text{O})\text{NEt}_2$	2.59	2.43	0.0373	11.15	0.6
32	Ethyl formate	$\text{HCOOC}_2\text{H}_5$	2.24	2.10	0.0277	12.53	0.6
33	Hexamethylphosphoramide	$[(\text{CH}_3)_2\text{N}]_3\text{PO}$	2.52	2.29	0.0296	11.13	0.6
34	Hexyl acetate	$\text{CH}_3\text{C}(\text{O})\text{Ohex}$	1.94	1.78	0.0138	12.28	0.4
35	Methyl acetate	$\text{CH}_3\text{C}(\text{O})\text{OCH}_3$	2.35 <sup>a</sup>	2.06	0.0260	12.50	0.2
36	<i>N,N</i> -Dimethylacetamide	$\text{CH}_3\text{CON}(\text{CH}_3)_2$	2.70 <sup>a</sup>	2.61	0.0439	10.93	1
37	<i>N,N</i> -Dimethylcyanamide	$(\text{CH}_3)_2\text{NCN}$	2.81 <sup>a</sup>	2.79	0.0531	11.52	0.2
38	<i>N,N</i> -Dimethylformamide	$\text{HCON}(\text{CH}_3)_2$	2.80	2.87	0.0550	11.07	1
39	<i>N</i> -Methylimidazole	$\text{C}_4\text{H}_6\text{N}_2$	2.60	2.92	0.0531	10.04	0.8
40	<i>N</i> -Methylpyrrolidinone	$\text{CH}_2\text{CH}_2\text{CH}_2\text{CONCH}_3$	2.62	2.55	0.0411	10.92	0.6
41	<i>n</i> -Butyronitrile	$\text{CH}_3\text{CH}_2\text{CH}_2\text{CN}$	2.70	2.40	0.0417	13.17	0.6
42	<i>n</i> -Decane	$\text{C}_{10}\text{H}_{22}$	0.90	1.32	0.0002	14.58	0.2
43	<i>n</i> -Heptane	$\text{C}_7\text{H}_{16}$	0.79	1.31	0.0000	14.75	0.2
44	<i>n</i> -Hexane	$\text{C}_6\text{H}_{14}$	0.68	1.31	0.0000	14.81	0.2
45	<i>n</i> -Nonane	$\text{C}_9\text{H}_{20}$	0.90	1.31	0.0000	14.66	0.2
46	<i>n</i> -Pentane	$\text{C}_5\text{H}_{12}$	0.57	1.30	0.0001	14.89	0
47	Nitrobenzene	$\text{C}_6\text{H}_5\text{NO}_2$	2.61	3.01	0.0539	9.36	1
48	Nitroethane	$\text{C}_2\text{H}_5\text{NO}_2$	2.78 <sup>a</sup>	3.09	0.0666	11.70	0.2
49	Nitromethane	$\text{CH}_3\text{NO}_2$	3.07	3.44	0.0812	11.85	0.8
50	Tetrahydropyran	$(\text{CH}_2)_5\text{O}$	1.98	1.81	0.0179	13.37	0.4
51	Propionitrile	$\text{C}_3\text{H}_5\text{CN}$	2.80	2.62	0.0519	13.50	0.6
52	Propyl acetate	$\text{CH}_3\text{C}(\text{O})\text{OPr}$	2.05	1.93	0.0200	12.31	0.4
53	Propylene carbonate	$(\text{CH}_2)_3(\text{O})_2\text{CO}$	3.10	3.24	0.0757	12.77	0.6
54	Pyridine	$\text{C}_5\text{H}_5\text{N}$	2.44	2.29	0.0272	10.07	1
55	Quinoline	$\text{C}_8\text{H}_7\text{N}$	2.30	2.20	0.0168	8.70	0.2
56	Tetrahydrofuran	$(\text{CH}_2)_4\text{O}$	2.08	2.02	0.0266	13.32	1
57	Tetrahydrothiophene	$(\text{CH}_2)_4\text{S}$	1.99 <sup>a</sup>	2.35	0.0269	9.48	0.2
58	Tetramethylene sulfone	$(\text{CH}_2)_4\text{SO}_2$	2.88 <sup>a</sup>	2.72	0.0489	11.10	0.4
59	Tetramethylurea	$[(\text{CH}_3)_2\text{N}]_2\text{CO}$	2.48 <sup>a</sup>	2.25	0.0287	10.94	0.6
60	Thiophene	$(\text{CH})_4\text{S}$	1.83 <sup>a</sup>	1.82	0.0048	9.46	0.2
61	Toluene	$\text{C}_6\text{H}_5\text{CH}_3$	1.66	1.73	0.0029	9.86	0.8
62	Tributyl phosphate	$(n\text{-C}_4\text{H}_9\text{O})_3\text{PO}$	2.30	1.74	0.0109	11.88	0.4
63	Trichloromethane	$\text{CCl}_3\text{H}$	1.74	1.92	0.0169	11.46	0
64	Triethylamine	$(\text{C}_2\text{H}_5)_3\text{N}$	1.43	1.69	0.0090	11.97	1
65	Triethyl phosphate	$(\text{C}_2\text{H}_5\text{O})_3\text{PO}$	2.55 <sup>a</sup>	1.88	0.0175	12.12	0.6
66	Trimethylbenzene	$\text{C}_6\text{H}_3(\text{CH}_3)_3$	1.54 <sup>c</sup>	1.69	0.0006	9.71	0.2
67	Trimethyl phosphate	$(\text{CH}_3\text{O})_3\text{PO}$	2.79 <sup>a</sup>	2.04	0.0245	12.26	0.2

$S'$  values are collected from four sources. <sup>a</sup> Ref. 9. <sup>b</sup> Ref. 5. <sup>c</sup> Ref. 10. <sup>d</sup> Weights are cited in ref. 5.

the *cis* conformation fits  $S'$  but not the experimental dipole moment (Table 5). Thus, the molecular descriptors that are generated from the boat conformation of 1,4-dioxane and the *cis* conformation of 1,1,2-trichloroethane reflect the rele-

vant molecular properties for the interaction of these molecules with polar solutes in solution. The chair form for dioxane, and *cis*-1,1,2-trichloroethane are included in the working set.

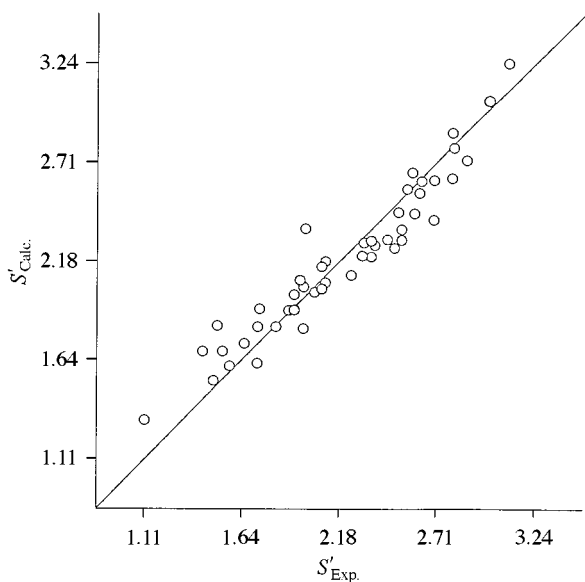


Fig. 4 The calculated vs. experimental  $S'$  of 48 diverse solvents using the two-parameter correlation eqn. (8)

Conformational problems must be considered for any relatively stable conformers (within 4 kcal mol<sup>-1</sup> to the global minimum) with higher dipole moments than the lowest energy conformer. Table 5 provides a collection of solvents, their dipole moments for different conformations, the  $S'$  values and the absolute energies relative to the global minimum. Butyl acetate, ethyl acetate, hexyl acetate, methyl acetate and propyl acetate have remarkably high calculated dipole moments for alternative conformations with energies within 4 kcal mol<sup>-1</sup> of the minimum energy. However, these solvents have large activation energies for rearrangement because C–O  $\pi$ -bonds are broken in the transition state. Using the higher dipole moments resulted in large deviations in the  $S'$  fit. We propose that in the lifetime of a solvent in the solute vicinity rearrangement of all these molecules does not occur. These solvents, labelled as *c* in Table 4, were excluded from the working set. The subsequent correlation for the remaining 48 solvents is then updated by eqn. (8) with  $R^2 = 0.9587$ ,  $F = 523$  and averaged absolute error

$$S' = (0.60 \pm 0.15) + (24.08 \pm 0.77)\mu/MV + (10.46 \pm 1.61)/E_{\text{HOMO-LUMO}} \quad (8)$$

of 0.11 (Fig. 4). The regression analysis was repeated for a set (comparing set) of 57 solvents using the experimental dipole moments and molar volumes,<sup>15,18</sup> along with the calculated reciprocal of HOMO–LUMO energy gap. With  $R^2$  equal to 0.9241 and  $F = 329$ , an average absolute error of 0.16 resulted. Using the experimental dipole causes 1,1,2-trichloroethane to be the most distinct outlier, as discussed before. Removing the outliers with predicted  $S'$  values exceeding 2.5 times the average absolute error, the correlation is refined giving eqn. (9) with

$$S' = (1.28 \pm 0.14) + (24.98 \pm 0.76)\mu/MV_{\text{exp}} + (3.49 \pm 1.53)/E_{\text{HOMO-LUMO}} \quad (9)$$

$R^2 = 0.9615$ ,  $F = 575$  and an average absolute error of 0.09. The statistical quality is similar to the results for fitting the calculated descriptors to experimental measurements adding confidence to this approach.

The successive regression coefficients for each descriptor, the respective standard errors, the  $R^2$  and  $F$  values and the t-test for eqns. (8) and (9) are provided in Table 6. The coefficients of these two descriptors are similar in all cases. The volume-weighted total dipole moment of the molecule and the reciprocal of the HOMO–LUMO energy gap account for most of

the changes in  $S'$  in both equations. Thus,  $S'$  is determined by the dipole density and the dispersion of the solvents.

Table 7 gives the predicted  $S'$  values calculated from eqn. (8). The calculated molecular volume weighted dipole moments and HOMO–LUMO energy gaps are also provided in Table 7. For the 15 reliable  $S'$  values weighted 1, the percent average absolute error is 5%. Weights were used in all the above fits to emphasize the best known parameters and not to dramatize the correlation coefficient. In fact, the unweighted correlation for the eqn. (8) data set gives eqn. (10) with  $R^2 = 0.9185$ ,  $F = 254$  and an average absolute error of 0.13.

$$S' = (0.67 \pm 0.18) + (25.77 \pm 1.15)\mu/MV + (9.30 \pm 1.77)/E_{\text{HOMO-LUMO}} \quad (10)$$

## Conclusions

The MQSPR approach can be employed to estimate the  $S'$  parameter of the unified, non-specific solvent polarity scale for new molecules. The dipole density  $\mu/MV_{\text{cal}}$  of the molecule and the reciprocal of the HOMO–LUMO energy gap are used successfully in the two-parameter correlation given by eqn. (8). These two descriptors are totally orthogonal. The correlation equation not only provides estimates of  $S'$  but does so by employing a model that is consistent with recognized dispersion and polarity contributions to non-specific solvation. The correlation also provides quantum-chemically calculable criteria for solvent selection, and enables one to analyse conformation dependent solvation influences.

The MQSPR approach is not restricted to solvation and is recommended as an alternative to the subset regression of the multiple QSPR descriptors for determining the intrinsic dimensions of other physicochemical properties. Model based predictions can be used with greater confidence on new systems than correlations to parameters that have no apparent meaning in the context of the properties predicted.

## Acknowledgements

The authors acknowledge the helpful discussions and encouragement provided by Professor A. R. Katritzky of this department.

## References

- (a) J. G. Kirkwood, *J. Chem. Phys.*, 1934, **2**, 351; (b) L. Onsager, *J. Am. Chem. Soc.*, 1936, **58**, 1486; (c) J. G. Kirkwood, *J. Chem. Phys.*, 1939, **7**, 911; (d) H. Block and S. M. Walker, *Chem. Phys. Lett.*, 1973, **19**, 363; (e) J. E. Brady and P. W. Carr, *J. Phys. Chem.*, 1985, **89**, 5759.
- R. S. Drago, *J. Chem. Soc., Perkin Trans. 2*, 1992, 1827 and references therein.
- (a) M. J. Kamlet, R. W. Taft and J.-J. M. Abboud, *J. Am. Chem. Soc.*, 1977, **99**, 8325; (b) M. J. Kamlet, J. L. M. Abboud and R. W. Taft, *Prog. Phys. Org. Chem.*, 1981, **13**, 485; (c) M. J. Kamlet and R. W. Taft, *Prog. Org. Chem.*, 1983, **48**, 2877; (d) M. J. Kamlet, R. M. Doherty, M. H. Abraham and R. W. Taft, *Quant. Struct.-Act. Relat.*, 1988, **7**, 71; (e) C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry*, 2nd edn., VCH, Weinheim, 1988; (f) C. Reichardt, *Chem. Rev.*, 1994, **94**, 2319.
- (a) R. S. Drago, *J. Org. Chem.*, 1992, **57**, 6547; (b) R. S. Drago, M. S. Hirsch, D. C. Ferris and C. W. Chronister, *J. Chem. Soc., Perkin Trans. 2*, 1994, 219; (c) D. C. Ferris and R. S. Drago, *J. Am. Chem. Soc.*, 1994, **116**, 7509; (d) R. S. Drago, D. Kovala-Demertzi and D. C. Ferris, *J. Coord. Chem.*, 1994, **32**, 145; (e) R. S. Drago and D. C. Ferris, *J. Phys. Chem.*, 1995, **99**, 6563.
- P. Bustamante and R. S. Drago, *J. Phys. Chem. B*, 1997, **101**, 5002.
- (a) G. R. Famini and L. Y. Wilson, *J. Phys. Org. Chem.*, 1992, **5**, 395; (b) G. R. Famini, W. P. Ashman, A. P. Mickiewicz and L. Y. Wilson, *Quant. Struct.-Act. Relat.*, 1992, **11**, 162; (c) G. R. Famini, B. C. Marquez and L. Y. Wilson, *J. Chem. Soc., Perkin Trans. 2*, 1993, 773; (d) G. R. Famini and L. Y. Wilson, *J. Chem. Soc., Perkin Trans. 2*, 1994, 1641; (e) A. H. Lowrey, C. J. Cramer, J. J. Urban and G. R. Famini, *Comput. Chem.*, 1995, **19**, 209; (f) C. J. Cramer, G. R.



- Famini and A. H. Lowery, *Acc. Chem. Res.*, 1993, **26**, 599; (d) H. D. William and G. R. Famini, *J. Chem. Soc., Perkin Trans. 2*, 1996, 83.
- 7 A. R. Katritzky, V. S. Lobanov and M. Karelson, *CODESSA Training Manual*, 1995; *CODESSA Reference Manual*, 1994.
- 8 A. R. Katritzky, L. Mu and M. Karelson, *J. Chem. Info. Comput. Sci.*, 1997, **37**, 156.
- 9 R. S. Drago, *Applications of Electrostatic-Covalent Models in Chemistry*, Surfside Scientific Publishers, Gainesville, FL, 1994.
- 10 J. E. George and R. S. Drago, *Inorg. Chem.*, 1996, **35**, 239.
- 11 *PCMODEL*, 5th Edn., Serena Software, Bloomington, IN, 1992.
- 12 M. M. Karelson, T. Tamm, A. R. Katritzky, S. J. Cato and M. C. Zerner, *Tetrahedron Comput. Methodol.*, 1989, **2**, 295.
- 13 L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Wiley, New York, 1986.
- 14 J. A. Paez, N. Campillo and J. Elguero, *Gazz. Chim. Ital.*, 1996, **126**, 307.
- 15 (a) Design Institute for Physical Property Data (DIPPR), *Physical and Thermodynamic Properties of Pure Chemicals: Data Compilation*; ed. T. E. Daubert and R. P. Danner, Hemisphere Publishing, New York, 1989, vols. 1–4; (b) *Handbook of Chemistry and Physics*, 71st edn., ed. R. C. Weast, CRC Press, Cleveland, Ohio, 1990.
- 16 (a) A. T. Amos and B. L. Burrows, *Adv. Quantum Chem.*, 1973, **7**, 289; (b) N. Rosch and M. C. Zerner, *J. Phys. Chem.*, 1994, **98**, 5817.
- 17 (a) A. R. Katritzky, L. Mu and M. Karelson, *J. Chem. Inf. Comput. Sci.*, 1996, **36**, 1162; (b) A. R. Katritzky, L. Mu, V. S. Lobanov, M. Karelson, *J. Phys. Chem.*, 1996, **100**, 10 400.
- 18 (a) A. L. McClellan, *Tables of Experimental Dipole Moments*, W. H. Freeman and Company, California, 1963; (b) A. L. McClellan, *Tables of Experimental Dipole Moments*, vol. 2, Rahara Enterprises, California, 1973; (c) A. L. McClellan, *Tables of Experimental Dipole Moments*, vol. 3, Rahara Enterprises, California, 1989.

Paper 7/02584J

Received 15th April 1997

Accepted 15th September 1997