The Electrostatic Origin of Abraham's Solute Polarity Parameter

J. Samuel Arey*

Department of Marine Chemistry and Geochemistry, MS #4, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

William H. Green, Jr.

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Philip M. Gschwend

Ralph M. Parsons Laboratory, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received: December 2, 2004; In Final Form: February 14, 2005

A computational method was developed which relates the empirical linear solvation energy relationship (LSER) solute polarity parameter, S (formerly denoted π_2^H), to two more fundamental quantities: a polarizability term and a computed solvent-accessible-surface electrostatic term. Electrostatics computations were conducted explicitly or with dielectric field polarizable continuum models (PCM, SCIPCM, IPCM), employing a density functional theory (B3LYP/6-311G(2df,2p)) or efficient Hartree−Fock (HF/MIDI!) method for 90 polar and nonpolar organic solutes. Electrostatic parameters calculated at electron isodensity solute surfaces were found to produce significantly better correlations with empirical S values than the same electrostatic parameters deduced from a fixed Bondi atomic radii based surface. The best-fit expression was found employing SCIPCM/IPCM at the 0.0004 e^-/bohr^3 solvent-accessible-surface: $S_{fit} = 0.46E - 0.091\Sigma V_s^2$, with squared correlation coefficient = 0.96 and standard deviation = 0.10, where E is a measured solute excess polarizability scale and ΣV_s^2 is a quantum-calculated solute electrostatic descriptor in kcal Å/mol. The resulting model is more accurate than previously developed estimation approaches and relies on only two fitted coefficients; it has the potential advantage of applicability to any solute composed of C, H, N, O, S, F, Cl, and Br. Finally, this investigation offers quantitative insight into the relative contributions of solute polarity and solute polarizability to the empirical LSER polarity parameter, S.

Introduction

The development of linear solvation energy relationships (LSERs) has contributed significant insight into the physical chemical processes governing solute−solvent interactions.[1] LSERs have been shown to accurately predict solvation free energies for a wide range of dilute solutes across different solvent environments.[2−4] Consequently, LSERs have potential applicability in diverse fields including separation sciences, environmental and chemical engineering, toxicology, and pharmacology, among others. Despite these successes and efforts to explain some LSER terms using proposed quantum-calculated[5] or thermodynamic properties,[6] the determination of most LSER parameter values remains essentially empirical. In particular, the solute polarity scales require a considerable amount of experimental data to fit,[7,8] and they continue to elude satisfactory correlations with more fundamental quantities. To date, the most extensively developed empirical solute polarity parameter, S, is that of Abraham and co-workers. The main aim of this work was to develop a method for estimating S values using molecular orbital calculations, which would in turn lead to accurate solvation energy estimates for unstudied solutes in many liquid−liquid and liquid−gas systems. Additionally, such an investigation could shed light on the physical origins of this free energy parameter.

A physical understanding of S must be placed in context of the development of LSERs. The LSER equation formulated by Abraham and co-workers is[3]

$$\log P = vV + eE + sS + aA + bB + c \qquad (1)$$

where P is a partitioning property of a solute between two bulk phases of interest, and v, e, s, a, b, and c are fitted coefficients, characteristic of a given two-phase system. The reader should note that the solute parameters V, E, S, A, and B are usually referred to as V_x, R_2, π_2^H, $\Sigma\alpha_2^H$, and $\Sigma\beta_2^H$ respectively, in older notation. V is a group-contributable solute volume which accounts for both the solvent cavitation energy and part of the solute−solvent London dispersion interaction (which increases with solute size).[9] E is the "excess molar refraction" of a solute: the measured liquid or gas molar refraction at 20 °C minus that of a hypothetical alkane of identical volume.[10] For compounds which are solids at 20 °C, group additivity or other estimation methods may be used to determine E.[11] The eE term is intended to capture solute−solvent interactions which involve an induced dipole (polarization) on the solute beyond what is accounted for by the vV term. E and V are therefore independently derived and clearly physically interpretable, unlike the remaining solute parameters in eq 1. A and B refer to the total hydrogen bond donating and hydrogen bond accepting capacities of the solute, respectively. The polarity/polarizability parameter, S, is believed to reflect the interactions associated with both induced and stable polarity on the solute. Finally, the fitted regression constant term, c, depends on the standard states and units of the partitioning property, P. Goss has additionally proposed that, in the case where P is a gas−liquid partition coefficient in molar concentration units, c may contain an

* Corresponding author. Telephone: (508) 289-3551. E-mail: arey@alum.mit.edu.

10.1021/jp044525f CCC: $30.25
Published on Web 03/18/2005

"accessible free solvent volume" entropy term.[6] Physical explanation of *A*, *B*, and *S* relies on the assumed linear separability of the underlying processes that they are intended to represent. To the extent that this assumption is invalid, the LSER parameters must reflect some blending of processes but may nevertheless give mathematically convenient results. Hence, to understand what is known about the *S* scale, we first examine its evolution among the concurrently developed LSER parameters.

The Original Development of *S*. The parameters *A*, *B*, and *S* represent "updated" parameters from a previous set of solvatochromic trial descriptors, α_2^H,[12] β_2^H,[13] and π^*.[7] These parameters have previously been related to solvation properties by fitting the coefficients of the equation:

$$\log P = l \log L^{16} + eE + s\pi^* + a\alpha_2^H + b\beta_2^H + c \quad (2)$$

where L^{16} is the air-hexadecane partition coefficient.[10] The development of these original parameters must be briefly reviewed in order to understand the basis of *S*. The hydrogen bonding parameters, α_2^H and β_2^H, were linear free energy scales of 1:1 hydrogen-bond complexation equilibrium constants in tetrachloromethane solvent. Conversely, the original polarity/polarizability parameter of Kamlet and co-workers, π^*, was a scale of solvent-induced spectral (frequency) shifts of the electronic transitions ($p \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$) of characteristic solutes which were not believed to engage in significant hydrogen-bonding with the selected solvents.[1,7,14] Although π^* was used as a solute descriptor, it actually reflected a *solvent* property: a measure of a solvent's ability to alter the spectral transitions of a characteristic set of solutes. With these assumptions, *s* values could be calibrated as two-phase system coefficients, taking π^* to be a solute parameter. Subsequently, Abraham et al. drew on these developments by proposing a LSER of gas chromatography retention times for several hundred solutes on 75 non-hydrogen-bond donating stationary phases (so that $b = 0$) using:

$$\log V^0 = l \log L^{16} + eE + s\pi^* + a\alpha_2^H + c \quad (3)$$

where V^0 is a retention capacity.[8] This resulted in a set of fitted coefficients (*l*, *e*, *s*, *a*, and *c*) obtained via multiple linear regression suited to characterizing diverse stationary phases. These workers then kept the stationary phase coefficients fixed and used the same set of data to *reverse fit* the polarity/polarizability and hydrogen-bond donating scales of the solutes, thereby producing an updated set of solute parameters, *S* and *A* (the $l \log L^{16}$ and *eE* terms were first subtracted from the dependent variable). Having determined *S* and *A*, values for *B* were similarly obtained, as follows. After using β_2^H as a trial descriptor to parametrize 16 water-organic solvent system LSER coefficients, these coefficients were fixed in order to isolate *B* in a reverse fit, with the contributions of other terms first subtracted from the dependent variable.[15]

Abraham and co-workers rationalized these reverse fit updates of the hydrogen bonding and polarity/polarizability descriptors as a way of "correcting" 1:1 complex solute parameters to reflect a more realistic set of interactions of the solute with multiple solvent molecules. It is difficult to assess what physical meaning may have been inserted into the parameters as a result of the updating procedure. For example, the *A* and *B* scales typically differed little, if at all, from the trial descriptors, α_2^H and β_2^H, for most solutes. This paved the way for characterizing new solutes, since *S*, *A*, and *B* values could subsequently be fitted via reverse regression from a large set of retention data in gas or liquid chromatography systems which had established system coefficients (i.e., *v*, *e*, *s*, *a*, *b*, and *c* values).[16] A history of this LSER approach and comparison to other LSER methods is discussed in a recent comprehensive review.[11]

Previous Correlations of the Polarity Parameters with Other Descriptors. Despite its somewhat complicated history, one may suspect that *S* is strongly related to both solute polarizability and solute charge density on the solvent-accessible-surface (hereafter referred to as "SAS"). Further, recognizing that Abraham and co-workers specifically attempted to separate the hydrogen-bonding and polarity contributions to the solute partial free energy of solvation, ΔG_{solv}, *S* may still reflect some mixing or interference with the hydrogen-bonding terms. Several workers have attempted to correlate the polarity scales π^* and *S* with theoretically conceived or calculated quantites. Brink et al. observed that "local charge separation" on the SAS of a solute may more accurately describe the solute's ability to create electrostatic interactions than do its dipole or multipole moments.[17,18] For example, some symmetric molecules (e.g., carbon dioxide, para-dinitrobenzene) have net zero dipole moments, but they exhibit significant charge separation at the SAS which can interact with surrounding solvent molecules. Brink et al. found a limited correlation between the π^* scale and a calculated parameter, Π, defined as the area-normalized summation of local charge separation on an operationally designated SAS of the solute. Zissimos et al.[19] showed that *S* could be partly explained by a multiple linear fit of the five computed descriptors of the ab initio solvation model of Klamt and co-workers, COSMO−RS,[20] finding a squared correlation coefficient $r^2 = 0.78$ and standard deviation $\sigma = 0.22$ for a set of 470 neutral organic compounds. Other workers have investigated correlations between π^* or *S* and area-normalized summations of solute SAS charge or its square,[21] measured and calculated dipole moments,[22] and various related quantities including summed atomic charges, calculated HOMO−LUMO energy gaps, and topological indices.[23,24] Still other investigators have proposed mixing computed and empirically derived solute descriptors to generate revised LSERs.[5,25] Most successfully, group contribution approaches have also been applied to the problem of estimating *S*. Platts et al. developed a comprehensive group contribution method consisting of 81 functional fragments for predicting *S*, finding $r^2 = 0.92$ and $\sigma = 0.16$ for the regression set.[26] Abraham additionally showed that for several families of aromatic compounds, *S* can be accurately estimated from a regression including both solute dipole moment and empirically fitted group-contribution parameters.[16] Such group-additivity estimation methods are practical and useful, as long as the group values are available. However a more fundamentally based approach might allow estimates of *S* in cases where functional group values exhibit poor additivity or have not yet been defined. It should finally be noted that Weckwerth et al. initiated the development of a separate LSER system based on reference solutes, which they contended may provide "purer" solute descriptors.[27] From the accomplishments of these groups, we concluded that the Debye-type and Keesom-type contributions to *S* may be related to electrostatic properties at the solute SAS. Since *S* exhibits reasonable functional additivity, we hypothesized that a correlation which relies on SAS area-aggregated, rather than area-normalized, charge descriptors was appropriate. Additionally, because *S* is partly rooted in dispersion interactions,[7,16] we considered it useful to separate out a partial dependence of *S* on solute polarizability (measured directly using the index of refraction).

On the basis of these considerations, the goals of this work were as follows: (1) to relate *S* with a computed solute electrostatic component and the solute excess polarizability scale, *E*; (2) to attempt to rationalize the contributions of these more fundamental quantities to the *sS* solvation free energy term of a LSER. We therefore present a methodology for calculating an appropriate solute electrostatic descriptor. We then discuss the extent to which the computed electrostatic term and *E* can explain *S* for a diverse set of solutes. Finally, we compare these results to some other approaches which have been previously suggested for calculating either *S* or π^*.

Method

Development of a Computed Electrostatic Descriptor. Any computational estimate of *S* should incorporate the simplifying assumptions inherent in the LSER formulation. Most notably, LSERs presuppose that the physical processes governing the solute are mathematically separable from those governing the solvent. For example, a *S* value is considered a constant property of the solute, regardless of the solvent. All information about the solvent relating to *S* is reflected by the LSER multiplying coefficient, *s*. This approximation implies that a computed analogue of *S* should be linear with respect to any solvent properties that may enter the calculation; in fact, it would be difficult to justify otherwise. Fortunately, classical electrostatic theory offers such an approach.

In classical electrostatics, the interaction energy, ΔU_q, between a solute, 1, and the surrounding solvent medium, 2, can be described as follows:[28]

$$\Delta U_q = \frac{1}{2}\int_{all\ space} \phi_1(\mathbf{r})\rho_2(\mathbf{r})\ \mathrm{d}\tau \tag{4}$$

Here $\rho_2(\mathbf{r})$ corresponds to the charge density of the solvent and $\phi_1(\mathbf{r})$ is the solute electrostatic potential at any point, $\mathbf{r}$, in infinitesimal volume, $\mathrm{d}\tau$. To facilitate the evaluation of ΔU_q, it is frequently assumed that a meaningful solute−solvent boundary may be reliably defined (although this is a disputed concept[29]). In this case, eq 4 is recast as an integral over so-called "virtual" surface charges at the SAS of the solute, $\sigma_2(\mathbf{r})$, which reflect the solvent's electrostatic interaction with the solute:[30]

$$\Delta U_q = \frac{1}{2}\oint_{SAS} \phi_1(\mathbf{r})\sigma_2(\mathbf{r})\ \mathrm{d}A \tag{5}$$

Solvent charge density at the solute−solvent boundary (the SAS) may be approximated by treating the solvent as a continuous linear dielectric medium.[30] In other words, $\sigma_2(\mathbf{r})$ "mirrors" the electrostatic field of the solute with a proportional field damped by the solvent dielectric constant, ϵ_2. Consequently, the solvent charge at the solute surface may be related to the normal component of the solute electrostatic field, $D_{1,n}(\mathbf{r})$, at the solute−solvent boundary as follows:[28]

$$\frac{\sigma_2(\mathbf{r})}{\epsilon_o} = \left(\frac{1}{\epsilon_2} - 1\right) D_{1,n}(\mathbf{r}) \tag{6}$$

where ϵ_2 is a constant scalar if the dielectric medium is assumed homogeneous and isotropic, and ϵ_o is the dielectric constant of a vacuum. The validity of linear solvent response for nonionic organic solvents has been evaluated in various Monte Carlo and molecular dynamics simulations.[31−34] These studies suggest that the linear solvent response approximation is most accurate when solute polarity exceeds that of the solvent. Linear solvent response may nevertheless provide useful estimates of electrostatic effects for both semipolar and polar solutes, since larger relative errors in ΔU_q are tolerable in cases where electrostatic effects make a smaller contribution to the total solvation partial free energy. The linear response approximation additionally implies that the electrostatic component of the partial free energy of solvation is given by the solute−solvent interaction energy.[31] Combining eqs 5 and 6,

$$\Delta U_q = \frac{1}{2}\left(\frac{1}{\epsilon_2} - 1\right)\epsilon_o \oint_{SAS} \phi_1(\mathbf{r}) D_{1,n}(\mathbf{r})\ \mathrm{d}A\ [\mathrm{kcal/mol}] \tag{7}$$

ΔU_q can be expressed directly in terms of the solute properties, $\phi_1(\mathbf{r})$ and $D_{1,n}(\mathbf{r})$. This computed electrostatic scale therefore suits the separability assumption of the corresponding LSER term, *S*. The computational realization of ΔU_q could be carried out in different ways, depending on what further assumptions are used (as described in the subsequent section).

The *S* parameter is believed to include both electrostatic effects and some solute polarization information; therefore we proposed that a linear combination of the measured solute excess polarizability and the computed electrostatic energy may explain *S*, to first order:

$$S_{fit} = \lambda_e E + \lambda_q \Delta U_q \tag{8}$$

where λ_e and λ_q are characteristic coefficients, optimized via multiple linear regression using literature (measured) *E* and *S* values and our computed ΔU_q values to produce S_{fit} values for 90 organic solutes.

Molecular Orbital Computations of ΔU_q. All solute geometry optimizations and electrostatic energy computations were performed using Gaussian98,[35] with "tight" SCF (self-consistent field) convergence criteria for the wave function computation. 90 solutes were optimized to energetically minimized nuclear geometries using both (1) the hybrid HF-DFT method B3LYP[36] with the 6-31G(d,p) basis set and (2) the HF method using the smaller MIDI! basis set.[37] The popular B3LYP method was chosen because it has been found to predict geometries, energetics, and electrostatic interactions more accurately than some other DFT and ab initio methods,[38,39] and the MIDI! basis set was employed because it has been optimized specifically for charge-property calculations. Since the role of *S* is most prominent in the presence of strongly polar solvents, all geometry optimizations utilized a dielectric continuum field corresponding to aqueous solution ($\epsilon_2 = 78.3$) using the polarizable continuum model (PCM[40,41]). Several single point electrostatic energy calculations were performed, using different combinations of the (1) B3LYP/6-31G(d,p) or HF/MIDI! optimized solute geometries and (2) SAS virtual solvent charges, $\sigma_2(\mathbf{r})$, from either the polarizable continuum model or the self-consistent isodensity polarizable continuum model (again assuming $\epsilon_2 = 78.3$). In single point computations, eq 7 was integrated over either a fixed-atomic Bondi radii surface (ΔU_q^B) or a solute electron isodensity surface (denoted ΔU_q^I), described as follows (Table 1).

ΔU_q^B [kcal/mol] values were computed in the presence of PCM-computed solvent charges, using both the B3LYP/6-311G-(2df,2p) and HF/MIDI! methods. The polarizable continuum model (PCM)[40,41] is a widely used solvation model which approximates $\sigma_2(\mathbf{r})$ as a set of discrete charges at the SAS_B, where SAS_B is defined as the outer surface carved by Bondi atomic radii[42] multiplied by 1.2. The SAS_B charges act to stabilize and distend the solute wave function. The PCM also incorporates the polarization response of SAS_B charges to each

TABLE 1: Methods Used to Compute the Solute Electrostatic Descriptor (ΔU_q in kcal/mol with ΣV_s^2 in kcal Å/mol)

electrostatic descriptor	solvent charges[a]	single point method	SAS used for eq 7
ΔU_q^B	PCM	B3LYP/6-311G(2df,2p)	1.2 Bondi radii
ΔU_q^B	PCM	HF/MIDI!	1.2 Bondi radii
ΔU_q^I	SCIPCM	HF/MIDI!	0.0004 e$^-$/bohr^3
ΔU_q^I	PCM	B3LYP/6-311G(2df,2p)	0.0004 e$^-$/bohr^3
ΔU_q^I	PCM	HF/MIDI!	0.0004 e$^-$/bohr^3
ΔU_q^I	PCM	HF/MIDI!	0.0001 e$^-$/bohr^3
ΣV_s^2	SCIPCM	HF/MIDI!	0.0004 e$^-$/bohr^3
ΣV_s^2	PCM	B3LYP/6-311G(2df,2p)	0.0004 e$^-$/bohr^3
ΣV_s^2	PCM	HF/MIDI!	0.0004 e$^-$/bohr^3
ΣV_s^2	PCM	HF/MIDI!	0.0001 e$^-$/bohr^3

[a] Using $\epsilon_2 = 78.3$.

other (i.e., "self-polarization" of the surface charges) and includes $D_1(\mathbf{r})$ corrections for SAS_B curvature. Using these approximations, the PCM self-consistently calculates SAS_B charges, polarizes the solute electronic wave function in response, and (optionally) relaxes the solute geometry[41] in the dielectric bath. Notably, the PCM and most related models do not include corrections for solute−solvent hydrogen-bonding effects. Since continuum solvation models were designed for prediction of solvation energies in a variety of systems, neglect of hydrogen-bonding has usually been considered a shortcoming of the approach.[43] However, predictions of the *S* parameter may be well suited by the continuum model approximations, since *S* was intentionally designed to be independent of hydrogen-bonding effects.

It was desirable to evaluate the electrostatic energy at an electron isodensity SAS to allow comparisons with electrostatics computed using the fixed atomic radii surface (SAS_B). ΔU_q^I was calculated in the presence of SCIPCM or IPCM solvent charges of Foresman et al.,[44] using the HF/MIDI! method as follows. The solute wave function was first relaxed in a dielectric bath using the self-consistent isodensity polarizable continuum model (SCIPCM[44]). For cases in which the SCIPCM did not converge (10 out of 90 solutes), the isodensity polarizable continuum model (IPCM)[44] was applied. The SCIPCM and IPCM formulations of solute-dielectric field interactions are similar to that of the PCM as described previously; however the SCIPCM and IPCM place virtual solvent charges at a solute electron isodensity surface of 0.0004 e$^-$/bohr^3 (SAS_F). The SAS_F may be more favorable than the conventional SAS_B, because an isodensity surface reflects the extent of solvent access to the solute expected from electron cloud repulsions between molecules. Since the location of the SAS_F is itself a function of calculated SAS_F charges (unlike the fixed SAS_B), the SAS_F charges are incorporated into the solute Hamiltonian potential expression and the wave function is iteratively calculated until the charge updates converge. The IPCM computes SAS_F charges between SCF convergence cycles, whereas the SCIPCM embeds SAS_F charge computations directly into the SCF procedure. Although the SCIPCM or IPCM may provide a more realistic SAS than the PCM, there are practical disadvantages to their use. The SCIPCM and IPCM are more computationally expensive than the PCM[40] and they are less numerically stable than the PCM.[29] The resulting solute wave function was used to generate a fine grid discretized electron density (output with the Gaussian98 "cube" keyword). An isodensity 0.0004 e$^-$/bohr^3 surface (SAS_I) was numerically interpolated from the calculated grid of the electron density. The resulting surface had uniform 0.04 Å^2 resolution, corresponding to about 1700 surface elements for a single molecule of water. Subsequently, solute $D_1(\mathbf{r})$ and $\phi_1(\mathbf{r})$ values were found at the SAS_I element centers. The $D_1(\mathbf{r})$ and $\phi_1(\mathbf{r})$ values reflected only the solute wave function. In other words, the computed field and potential values at the SAS_I elements did not include the field and potential contribution of the SCIPCM or IPCM charges, although the solute wave function had been optimized in the SCIPCM or IPCM dielectric environment. The outward normal vector at each SAS_I element was approximated using the locations of several adjacent SAS_I element centers. Computed SAS_I normal vectors were then combined with $D_1(\mathbf{r})$ values to arrive at $D_{1,n}(\mathbf{r})$ estimates. ΔU_q^I could subsequently be numerically integrated over the SAS_I of each solute. Since the SCIPCM and IPCM were computationally expensive, it was desirable to also apply the PCM prior to integration of ΔU_q^I, in a new set of calculations using both the B3LYP/6-311G(2df,2p) and HF/MIDI! methods (Table 1). In other words, since application of the dielectric continuum was a separate step from integration of ΔU_q^I, it was possible to use the PCM to optimize the solute wave function in an aqueous dielectric ($\epsilon_2 = 78.3$) and subsequently integrate eq 7 at the SAS_I. Finally, in an additional set of calculations, ΔU_q^I was integrated over a fine-grid 0.0001 e$^-$/bohr^3 SAS_I in order to test the sensitivity of results to the isodensity surface location. HF/MIDI! was used for most of the sets of ΔU_q^I calculations because it was computationally expensive to use the SCIPCM or IPCM together with the B3LYP/6-311G(2df,2p) method for the largest solutes. Since SAS_I curvature corrections were not made for $D_{1,n}(\mathbf{r})$ estimates, the numerical integration of ΔU_q^I was additionally evaluated using several levels of SAS_I resolution with water as a test solute.

As outlined above, computation of $D_{1,n}(\mathbf{r})$ required numerical evaluation of the normal electrostatic field component at the solute SAS_I for a large number of points. This procedure added complexity to the method and may be susceptible to errors; hence it was desirable to generate a more tractable form of the integral in eq 7. As a proposed simplification, integration over the normal solute field was thus assumed proportional to integration over the solute potential at the solute SAS_I:

$$\oint_{SAS_I} \phi_1(\mathbf{r}) D_{1,n}(\mathbf{r}) dA = -\oint_{SAS_I} \phi_1(\mathbf{r}) \left[\frac{d\phi_1(\mathbf{r})}{dn(\mathbf{r})}\right] dA \propto -\oint_{SAS_I} \phi^2{}_1(\mathbf{r}) dA \quad (9)$$

where $D_{1,n}(\mathbf{r})$ is given by $-d\phi_1(\mathbf{r})/dn(\mathbf{r})$ (the gradient of the solute electrostatic potential along the SAS normal vector, $n(\mathbf{r})$), and $d\phi_1(\mathbf{r})/dn(\mathbf{r})$ is assumed proportional to $\phi_1(\mathbf{r})$. Equation 9 was evaluated by comparison of calculated $\phi_1(r)D_{1,n}(\mathbf{r})$ and $\phi^2{}_1(\mathbf{r})$ values over the entire set of SAS_I points on the 90 studied solutes. The validity of eq 9 is further discussed in the Results section. Equation 9 led to a new electrostatic scale proportional to ΔU_q^I which could now be defined (from eq 7) as

$$\Sigma V_s^2 \equiv -\frac{1}{2}\left(\frac{1}{\epsilon_2} - 1\right)\epsilon_o \oint_{SAS_I} \phi^2{}_1(\mathbf{r})\, dA \text{ [kcal Å/mol]} \quad (10)$$

In a new set of calculations, the validity of eq 9 was additionally judged by the observed correlation of ΣV_s^2 with ΔU_q^I for 90 solutes. Finally, regressions of eq 8 were also evaluated using ΣV_s^2 as a substitute for ΔU_q (Table 1).

It is important to note that computed ΔU_q values did not include the computed change in solute energy caused by polarization of the solute wave function in the dielectric field. It was considered advantageous to exclude PCM or SCIPCM/IPCM-induced solute polarization energies and independently fit the solute polarizability contribution (using *E*) in eq 8 for at least two reasons. First, the PCM and SCIPCM/IPCM-calculated solute polarization energies did not correlate well with the measured polarizability scale (data not shown). Second, the precise origin and magnitude of the solute polarizability contribution to *S* is unclear, since the LSER formulation already incorporates an *E* component in empirical fits of solvation data (as discussed in the Introduction). Investigators wishing to reproduce the computational method described herein should note that, by default, the PCM-computed ΔG_{solv} includes an estimated solvent cavitation energy term, an estimated solute−solvent dispersion−repulsion interaction energy term, a dielectric field-induced solute polarization energy term, and finally ΔU_q^B. These first three terms were not included in the analysis we conducted. Similarly, the default SCIPCM and IPCM-computed ΔG_{solv} output implicitly includes a solute polarization energy associated with the dielectric charges. Consequently, we decided to use the calculated electronic population to define an electron isodensity surface for explicit integration of eq 7 (as described previously), to provide a more direct estimate of ΔU_q^I.

Gas-Phase Dipole Moment Computations. As a validation of the computational accuracy of the molecular orbital computation methods for charge distribution properties, gas-phase dipole moments were computed and compared to experimental values available in the literature. The B3LYP/6-31G(d,p) and HF/MIDI! approaches were used to optimize the geometries of 45 solutes from the set of 90 considered for this study, in vacuo. The B3LYP/6-311G(2df,2p) and HF/MIDI! methods were then used to compute the dipole moments of these solutes (also in vacuo), and these calculated results were compared with measured gas-phase dipole moments.[45]

Selection of *S* Data. We selected solutes which exhibit a wide range of *S* values.[46,47] We included *n*-alkanes as reference compounds, since Abraham set $S_{alkane} = 0$.[8] A range of small to moderately sized aliphatic compounds (1−11 non-hydrogen atoms), often containing multiple moieties, composed the first subset of the list. A few homologous series were included to evaluate the effect of nonpolar chains. Aromatic compounds composed the second subset (ranging from 4 to 16 non-hydrogen atoms), some of which contained N, S, or O as ring members. Multiple moieties and some flexible chain substituents characterized many of the aromatic compounds. In both aliphatic and aromatic sets, a range of semipolar (e.g., olefin, amino, halide) to highly polar (e.g., sulfone, sulfoxide, amide, nitro) groups were tested. Additionally, in some cases one or more electron-withdrawing groups (such as halogens) were proximate to a polar group, causing important intramolecular interactions (e.g., 3-bromophenol and 2,2,2-trifluoroethanol). This range of compounds allowed us to evaluate the robustness of the model's applicability to small and moderately sized organic compounds containing C, H, N, O, S, F, Cl, and Br (only one compound containing P was included in the set).

Results and Discussion

Computation of Gas-Phase Dipole Moments. It was desirable to evaluate the reliability of the HF/MIDI! and B3LYP/6-311G(2df,2p) methods against an independently measurable molecular charge distribution property. Reported gas-phase dipole moments[48] compared favorably to those calculated in vacuo using both methods, with $r^2 = 0.96$ and rmse (root-mean-squared error) = 0.25 for HF/MIDI! computations (Figure 1) and $r^2 = 0.975$, rmse = 0.19 for B3LYP/6-311G(2df,2p) computations (not shown) for 45 compounds among the set of 90 considered in this work. It was reasonable to assume that fits of eq 8 would reflect this limitation in accuracy; that is, we did not expect to generate a *S* model of significantly better predictive quality than that found for gas-phase dipole moment calculations.

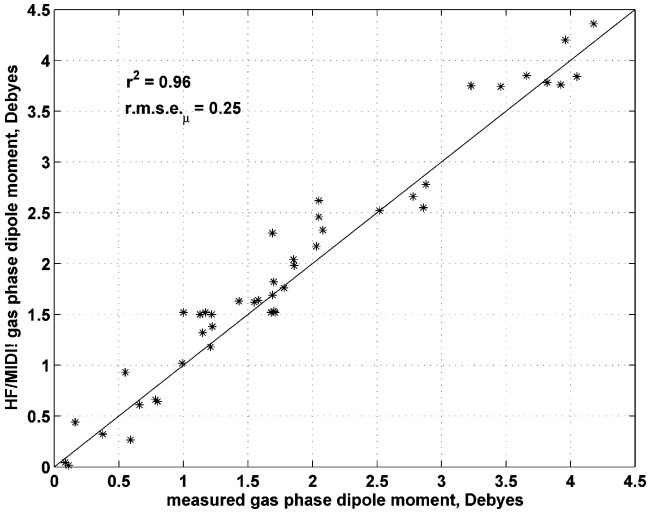

Figure 1. HF/MIDI! computed in vacuo dipole moments for 45 compounds.

***S* Regressions.** Regressions using 90 solutes (Table 2) showed accurate estimates of *S* using linear combinations of the computed electrostatics terms and measured solute polarizabilities (Table 3). HF/MIDI! electrostatic energies computed at the 0.0004 e^-/bohr^3 SAS_I in the presence of a SCIPCM/IPCM dielectric produced the best correlation with *S* values (Table 2):

$$S_{fit} = 0.48E - 0.105\Delta U_q^I \qquad (11)$$

$$r^2 = 0.94, \quad \sigma = 0.13$$

where ΔU_q^I is in kcal/mol. Using the HF/MIDI! method, S_{fit} was relatively insensitive to changes in the dielectric field method (PCM vs SCIPCM/IPCM) or location of the SAS_I (0.0004 vs 0.0001 e^-/bohr^3). This indicated that ΔU_q^I is probably a robust and physically meaningful parameter for *S*. By comparison, S_{fit} values calculated from electrostatic energies at the 120% Bondi radii solute SAS (ΔU_q^B) showed considerably weaker agreement with *S* ($r^2 < 0.80$ and $\sigma > 0.20$; Table 3). This suggested that an SAS based on electron isodensity, rather than the 120% Bondi radii SAS, is an appropriate physical surface for *S*.

In all regressions of eq 8 (Table 3), the measured excess polarizability and computed electrostatics term explained about $^1/_3$ and $^2/_3$ of S_{fit} variability, respectively, showing that both of these terms are important components of *S*. This indicated that while stable charge density at the solute surface generally dominates the *S* term, solute polarizability also contributes significantly. These results support the contention that Abraham and co-workers have indeed isolated a LSER term which quantitatively reflects mainly polarity and polarizability character of the solute.

A data-withholding test of each variant of eq 8 was conducted in order to evaluate its robustness for novel compounds not

TABLE 2: Regression Results for Eq 11 and Comparison of Calculated and Literature *S* Values (ΔU_q^I in kcal/mol)

solute	*E*	$-\Delta U_q^I$	calcd S_{fit}	measd *S*	solute	*E*	$-\Delta U_q^I$	calcd S_{fit}	measd *S*
propane	0.000	0.18	0.02	0.00	nitroethane	0.270	7.62	0.93	0.95
pentane	0.000	0.24	0.03	0.00	nitropropane	0.242	7.35	0.89	0.95
cyclohexane	0.305	0.22	0.17	0.10	ethyl acetate	0.106	4.77	0.55	0.62
1-hexene	0.078	0.83	0.12	0.08	acetic acid	0.265	7.29	0.89	0.65
propyne	0.183	2.27	0.33	0.25	2,2,2-trifluoroethanol	0.015	7.64	0.81	0.60
1-butyne	0.178	2.20	0.32	0.23	enflurane	−0.230	5.21	0.43	0.40
fluoromethane	0.066	3.29	0.38	0.35	isoflurane	−0.240	6.17	0.53	0.50
1-fluorobutane	0.017	2.95	0.32	0.35	trimethyl phosphate	0.113	11.28	1.24	1.10
1-fluoropentane	0.002	3.00	0.31	0.35	propionamide	0.440	9.21	1.18	1.30
tetrafluoromethane	−0.280	0.78	−0.05	−0.20	*N*-methylformamide	0.405	10.06	1.25	1.30
hexafluorosulfide	−0.600	0.32	−0.26	−0.20	dimethyl sulfone	0.590	14.45	1.80	1.70
chloroethane	0.227	2.82	0.41	0.40	*N*,*N*-dimethylacetamide	0.363	7.26	0.94	1.33
1-chlorobutane	0.210	2.81	0.40	0.40	*N*,*N*-dimethylformamide	0.367	8.05	1.02	1.31
1-chlorooctane	0.191	2.70	0.37	0.40	dimethyl sulfoxide	0.522	10.90	1.40	1.74
carbon tetrachloride	0.458	0.48	0.27	0.38	benzene	0.610	1.79	0.48	0.52
1,1,2-trichloroethane	0.499	4.08	0.67	0.68	toluene	0.601	1.72	0.47	0.52
hexachloroethane	0.680	0.65	0.40	0.22	ethylbenzene	0.613	1.71	0.48	0.51
bromoethane	0.366	2.75	0.46	0.40	naphthalene	1.340	2.64	0.92	0.92
1-bromobutane	0.360	2.72	0.46	0.40	phenanthrene	2.055	3.51	1.36	1.29
1-bromooctane	0.339	2.94	0.47	0.40	pyrene	2.808	3.83	1.76	1.71
dibromomethane	0.714	3.17	0.68	0.67	chlorobenzene	0.718	2.71	0.63	0.65
tribromomethane	0.974	2.49	0.73	0.68	1,2,4-trichlorobenzene	0.980	3.24	0.81	0.81
diethyl ether	0.041	1.72	0.20	0.25	1,2-dibromobenzene	1.190	3.31	0.92	0.96
dipropyl ether	0.008	1.84	0.20	0.25	aniline	0.955	4.71	0.96	0.96
tetrahydrofuran	0.289	2.70	0.42	0.52	*N*-methylaniline	0.948	3.74	0.85	0.90
carbon monoxide	0.000	0.69	0.07	0.00	*N*,*N*-dimethylaniline	0.957	2.84	0.76	0.84
carbon dioxide	0.150	3.12	0.40	0.42	phenol	0.805	5.91	1.01	0.89
carbon disulfide	0.877	0.26	0.45	0.21	*m*-cresol	0.820	5.66	0.99	0.87
dioxygen	0.000	0.07	0.01	0.00	benzyl alcohol	0.803	4.93	0.90	0.87
nitrous oxide	0.068	2.11	0.25	0.35	benzaldehyde	0.820	5.59	0.98	1.00
ethylamine	0.236	2.97	0.42	0.35	benzonitrile	0.742	6.91	1.08	1.11
propylamine	0.225	3.34	0.46	0.35	thiophene	0.687	2.27	0.57	0.56
butylamine	0.224	3.36	0.46	0.35	benzothiophene	1.323	3.00	0.95	0.88
water	0.000	6.96	0.73	0.45	thiazole	0.800	4.45	0.85	0.80
methanol	0.278	4.68	0.62	0.44	pyrazole	0.620	7.29	1.06	1.00
ethanol	0.246	4.22	0.56	0.42	benzophenone	1.447	5.56	1.28	1.50
1-propanol	0.236	4.23	0.56	0.42	4-cyanophenol	0.940	11.78	1.69	1.55
2-propanol	0.212	4.16	0.54	0.36	diethylphthalate	0.729	8.04	1.19	1.40
1-decanol	0.191	4.49	0.56	0.42	benzotrifluoride	0.225	3.26	0.45	0.48
acetone	0.179	5.46	0.66	0.70	3-bromophenol	1.060	6.77	1.22	1.15
butanone	0.166	4.78	0.58	0.70	benzamide	0.990	9.45	1.47	1.50
propanal	0.196	4.58	0.57	0.65	benzenesulfonamide	1.130	13.27	1.94	1.55
acetonitrile	0.237	7.13	0.86	0.90	methylphenyl sulfone	1.080	12.29	1.81	1.85
propionitrile	0.162	6.72	0.78	0.90	diphenyl sulfone	1.570	11.01	1.91	2.15
nitromethane	0.313	8.23	1.01	0.95	methylphenyl sulfoxide	1.080	9.81	1.55	1.85

TABLE 3: Regression Statistics for Eq 8 Using 90 Solutes (ΔU_q in kcal/mol; ΣV_s^2 in (kcal Å)/mol)

electrostatic descriptor	dielectric model[a]	method[b]	SAS used for electrostatics[c]	eq 8 regression best fit coeffs		regression statistics		data-withholding test statistics	
				$\lambda_e \pm \sigma_\lambda$ (weight)	$\lambda_q \pm \sigma_\lambda$ (weight)	r^2	σ	r^2	rmse
ΔU_q^B	PCM	B3LYP	1.2 Bondi radii	0.46 ± 0.22 (34%)	−0.095 ± 0.038 (66%)	0.72	0.28	0.70	0.28
ΔU_q^B	PCM	MIDI!	1.2 Bondi radii	0.45 ± 0.22 (31%)	−0.111 ± 0.041 (69%)	0.78	0.23	0.77	0.24
ΔU_q^I	SCIPCM	MIDI!	0.0004 au	0.48 ± 0.20 (33%)	−0.105 ± 0.035 (67%)	0.94	0.13	0.93	0.13
ΔU_q^I	PCM	B3LYP	0.0004 au	0.45 ± 0.21 (31%)	−0.124 ± 0.042 (69%)	0.92	0.14	0.91	0.15
ΔU_q^I	PCM	MIDI!	0.0004 au	0.48 ± 0.20 (33%)	−0.106 ± 0.036 (67%)	0.92	0.14	0.91	0.15
ΔU_q^I	PCM	MIDI!	0.0001 au	0.49 ± 0.20 (33%)	−0.153 ± 0.051 (67%)	0.94	0.13	0.93	0.13
ΣV_s^2	SCIPCM	MIDI!	0.0004 au	0.46 ± 0.20 (32%)	−0.091 ± 0.030 (68%)	0.96	0.10	0.95	0.11
ΣV_s^2	PCM	B3LYP	0.0004 au	0.44 ± 0.21 (31%)	−0.097 ± 0.033 (69%)	0.93	0.14	0.92	0.15
ΣV_s^2	PCM	MIDI!	0.0004 au	0.45 ± 0.20 (31%)	−0.094 ± 0.031 (69%)	0.95	0.11	0.94	0.12
ΣV_s^2	PCM	MIDI!	0.0001 au	0.47 ± 0.20 (32%)	−0.115 ± 0.038 (68%)	0.95	0.11	0.95	0.11

[a] Using $\epsilon_2 = 78.3$. [b] Indicates either a B3LYP/6-311G(2df,2p) or HF/MIDI! single point wave function computation. [c] 1 au = 1 e^-/bohr^3 for an isodensity SAS.

included in the regression set. This was done by determining the best-fit coefficients to eq 8 using 89 of the 90 solutes, and then using this regression to calculate S_{pred} for the withheld solute. This procedure was then repeated to generate 90 S_{pred} values that were independent of the regressions. All regressions and parameter statistics were calculated using singular value

decomposition.[49] Data-withholding tests suggested that the expected uncertainty of S_{pred} values for solutes outside of the regression sets were similar to regression statistics (Table 3). Hence, within the parameter space represented by our training set, eq 11 gave good predictions for new compounds.

Regression outliers for eq 8 using either of the computed ΔU_q^B or ΔU_q^I terms showed some systematic biases. The most egregious *S* overestimates were consistently strong hydrogen-bond donors (e.g., water, acetic acid, 2,2,2-trifluoroethanol, and benzenesulfonamide). Conversely, underestimated *S* outliers were usually strong hydrogen-bond acceptors (e.g., *N*,*N*-dimethylformamide, *N*,*N*-dimethylacetamide, dimethyl sulfoxide). This pattern of model prediction error could result from hydrogen-bonding interference during the development of *S* parameter values from data. As discussed in the Introduction, Abraham et al.'s development of *S* involved reverse fits which simultaneously updated both *A* and *S* values from trial descriptors. It is difficult to ascertain that these earlier investigations successfully removed all of the Debye and Keesom contributions to *A*. Or, particularly for highly polar solutes, blending of some of the *B* character into *S* values may have occurred, assuming that these contributions to the solvation free energy are in fact linearly separable. It was additionally useful to examine model performance for certain subsets of the data, since our chosen set of solutes is somewhat biased toward monofunctional compounds. If only the subset of 38 solutes containing two or more heteroatoms was considered, model deviation statistics using eq 11 were slightly worse ($r^2 = 0.92$ and $\sigma = 0.16$) than the statistics found for the entire regression set. Hence we suspect that model predictions for multifunctional, polar molecules such as carbohydrates may be worse than is suggested by the statistics shown here. Among the 31 solutes containing aromatic functional groups, model error statistics for eq 11 were comparable to those for the complete set: $r^2 = 0.92$ and $\sigma = 0.13$. Finally, we suspected that failure of the computational methods to evaluate properly surface potential on some highly polar moieties might explain some bias and error in *S* regressions. However, a comparison of HF/MIDI! calculated dipole moment residuals with S_{fit} residuals from eq 11 showed no correlation at all ($r^2 = 0.03$). Additionally, the accuracy of gas-phase dipole moment computations did not correlate with solute polarity (Figure 1). Consequently, we concluded that the computational methods were not responsible for the observed S_{fit} error bias toward compounds containing strong hydrogen-bonding moieties.

The best *S* regression overall was found with the HF/MIDI! computed ΣV_s^2 electrostatic descriptor (as opposed to a computed electrostatic energy, ΔU_q):

$$S_{fit} = 0.46E - 0.091\Sigma V_s^2 \qquad (12)$$

$$r^2 = 0.96, \quad \sigma = 0.10$$

employing the SCIPCM/IPCM at the 0.0004 e$^-$/bohr^3 SAS$_I$ (Figure 2), where ΣV_s^2 is in [kcal Å/mol]. In fact, substitution of the computed $D_{1,n}(\mathbf{r})$ by the $\phi_1(\mathbf{r})$ in SAS$_I$ integrals (eq 9) produced consistently improved correlations between the electrostatic descriptor and *S* (using HF/MIDI!, Table 3). This surprising result could be explained as a canceling of errors between eqs 9 and 11. Using the HF/MIDI! method with the

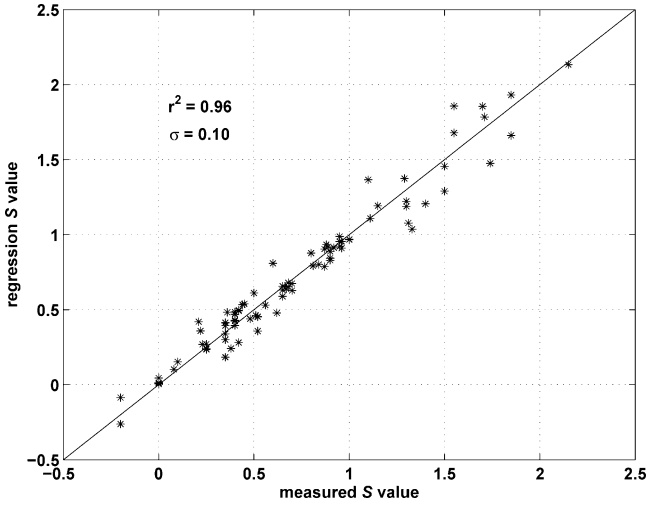

Figure 2. Plot of *S* regression results for eq 12.

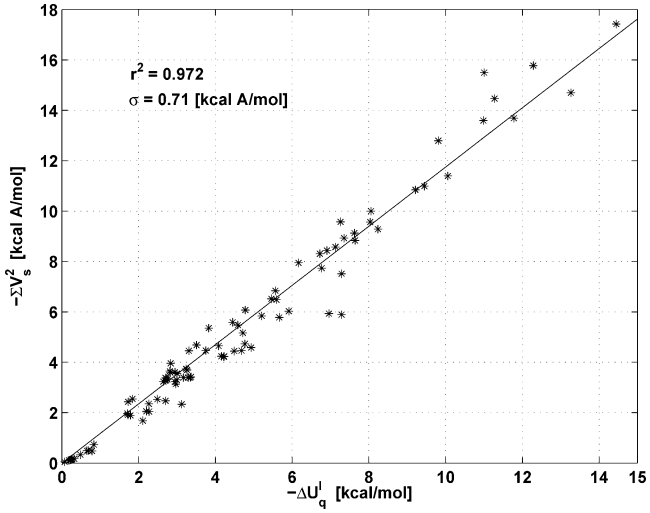

Figure 3. Correlation between ΔU_q^I and ΣV_s^2 for 90 solutes using the HF/MIDI! method with SCIPCM/IPCM at the 0.0004 e$^-$/bohr^3 SAS$_I$.

SCIPCM/IPCM at the 0.0004 e$^-$/bohr^3 SAS$_I$, it was found that (Figure 3):

$$\Sigma V_{s,fit}{}^2 = (1.17\ \text{Å})\Delta U_q^I\ [\text{kcal Å/mol}] \qquad (13)$$

$$r^2 = 0.97, \quad \sigma = 0.71\ [\text{kcal Å/mol}]$$

Solutes which had the most overestimated ΣV_s^2 values in this correlation were consistently strong hydrogen-bonding donors (e.g., water, acetic acid), whereas strong hydrogen-bond acceptors (e.g., diphenyl sulfone) were typically underestimated. The outlier bias of eq 13 therefore mirrors the outlier bias found for eq 11 using the HF/MIDI! method, creating offsetting errors. This explains how ΣV_s^2 was apparently the most successful electrostatic variable for predicting *S* values for the set of solutes considered here.

On Correlating Solute Electric Field with Electric Potential at the SAS. It would be reasonable to conclude that since eq 13 produces a good correlation ($r^2 = 0.97$), it is additionally the case that $\phi_1(\mathbf{r})D_{1,n}(\mathbf{r})$ correlates well with $\phi^2{}_1(\mathbf{r})$ locally at the SAS$_I$. However, using the HF/MIDI! SCIPCM/IPCM-computed 0.0004 e$^-$/bohr^3 SAS$_I$ surfaces on the 90 solute set, it was found that

$$\phi^2_{1,fit}(\mathbf{r}) = -(1.03\ \text{Å})\phi_1(\mathbf{r})D_{1,n}(\mathbf{r})\ [\text{V}^2] \qquad (14)$$

$$r^2 = 0.90$$

An apparent incongruity arises here: eq 13 suggests that $\phi^2{}_1(\mathbf{r})$ is a robust substitute for $\phi_1(\mathbf{r})D_{1,n}(\mathbf{r})$ in the SAS_I electrostatic energy integral, but in eq 14, $\phi^2{}_1(\mathbf{r})$ is a less accurate ($r^2 = 0.90$) explanatory variable for $\phi_1(\mathbf{r})D_{1,n}(\mathbf{r})$ at local points on the SAS_I. Additional analysis (data not shown) revealed that this contradiction could be explained as follows. Since eq 14 residuals are additive in the SAS_I electrostatic energy integral, they counterbalance in the integral to a significant extent. Additionally, it was found that molecules having a preponderance of positive $\phi^2{}_{1,fit}(\mathbf{r})$ residuals in eq 14 correlated with total molecular surface charge (ΔU_q^I). In other words, molecules having high ΔU_q^I values tended to have $\phi^2{}_1(\mathbf{r})$ values underpredicted by eq 14. These trends contributed information content to the regression between ΔU_q^I and ΣV_s^2, resulting in the robust correlation given by eq 13.

Comparison to Alternative Models. To date, the group contribution approach developed by Platts et al.[26] is the most comprehensively tested technique for *S* estimation. For the set of 90 solutes that we considered, 85 could be treated using the current incarnation of the method of Platts et al.,[50] since the software was unable to handle inorganic compounds or carbon monoxide. The error statistics for this solute set were $r^2 = 0.75$ and rmse = 0.25, suggesting that many compounds we have chosen pose a severe challenge to the group contribution approach. The worst outlier was carbon disulfide, having $S_{Platts} = 1.59$ and $S_{meas} = 0.26$.

We compared the results of eqs 11 and 12 to previous correlations that have been developed for π^* or *S*, using the set of compounds considered here. Lewis suggested a correlation of π^* with calculated dipole moments plus an intercept, finding a reasonable fit for 14 solutes ($r^2 = 0.91$).[22] Lamarche et al. suggested *S* fits with linear combinations of solute dipole moment, polarizability, and other quantities such as calculated atomic charges and HOMO−LUMO gap, finding correlations ranging from $r^2 = 0.76$ to $r^2 = 0.85$ for a set of 58 solutes. In this vein, we used the efficient HF/MIDI! method with a PCM dielectric field ($\epsilon_2 = 78.3$) to compute solute dipole moments, finding a correlation with *S* for the 90 solutes considered here

$$S_{fit} = 0.59E + 0.17\mu_{calc} \qquad (15)$$

$$r^2 = 0.85, \quad \sigma = 0.19$$

where the dipole moment is expressed in debye. A similar regression of *S* with $\mu_{calc}{}^2$ (which corresponds to the pairwise free energy of interaction between freely rotating dipoles[51]) yielded a comparable fit. Although convenient to compute, these regressions are biased against symmetric molecules, since such solutes may exhibit substantial solvent-accessible charge separation not reflected in their dipole moments (e.g., carbon dioxide, benzene). It is worth noting that although the solute dipole moment is commonly relied upon as an indicator of solute polarity in solvent environments, higher multipoles contribute significantly to the solute−solvent interaction energy. In fact, the marginal contributions of higher multipoles may be slowly convergent, and they may still be significant well beyond the 20th term in the multipole expansion.[29] Recognizing this deficiency of the dipole moment as an electrostatic descriptor, Brink et al. developed Π, an area-normalized summation of absolute elecrostatic surface potential:[17]

$$\Pi \equiv \frac{1}{A}\oint_{SAS}|\phi_1(\mathbf{r}) - \bar{\phi}_1|dA \qquad (16)$$

and an area-normalized summation of squared electrostatic surface potential, which they termed $\sigma_{tot}{}^2$.[18] Brink et al. found a limited correlation between π^* and Π plus a polarizability parameter plus an intercept (statistics were not given). Zou et al. recently improved this π^* correlation by including $\sigma_{tot}{}^2$ as an additional term ($r^2 = 0.93$ for 50 solutes).[21] Using HF/MIDI! calculations with the SCIPCM/IPCM dielectric field and a 0.0004 e^-/bohr^3 solute isosurface, we found the following correlation for *S* using the solute set presented here:

$$S_{fit} = 0.55E + 1.02\Pi \qquad (17)$$

$$r^2 = 0.80 \quad \text{and} \quad \sigma = 0.23$$

where Π is given in volts. Adding a $\sigma_{tot}{}^2$ term to the eq 17 regression undermined the statistical interpretability of the parameters and improved the fit little ($r^2 = 0.84$). Floating a constant failed to improve the correlation. The disparity in goodness of fit found between eq 17 and eq 11 is consistent with the notion that *S* reflects area-aggregated, rather than area-normalized, charge density on the solute surface.

Conclusions

A method has been developed to estimate the polarity/polarizability parameter, *S*, for new solutes. This empirical parameter has been purported to capture solute electrostatic contributions to the solvation free energy, with minimal interference from solute−solvent hydrogen-bonding interactions. Nevertheless, *S* has conventionally eluded reliable correlations with more fundamental quantities. Moreover, its ambiguous physical origin has been presumed to reflect a conserved solute property over a wide range of solvent environments. Despite its complicated inception, *S* appears to be accurately explained by two solute properties: a polarizability term and a computed solvent-accessible-surface electrostatic term. This result supports the contention that solute−solvent interaction free energies are mostly separable into solute−solvent hydrogen bonding, solvent cavitation, solute polarization, and solute−solvent electrostatic interactions. Additionally, correlations found between *S* and electrostatic descriptors consistently indicate that a solute electron isodensity surface is a better basis for electrostatic computations than a Bondi fixed atomic radii surface. Results here show that *S* is not very sensitive to the choice of isodensity surface in the 0.0001−0.0004 e^-/bohr^3 range, additionally corroborating the robustness of this particular type of surface. This directed the development of a model for *S*; however, it may additionally inform the ongoing debate over what type of solvent-accessible-surface is most appropriate for continuum solvation free energy computations more generally.

For practical applications of *S* estimation, we recommend eq 12, which has an estimated S_{pred} standard error of about 0.11; i.e., using the efficient HF/MIDI! method for computation of ΣV_s^2 at a 0.0004 e^-/bohr^3 SAS_I in the presence of a SCIPCM or IPCM dielectric field with $\epsilon_2 = 78.3$. However in (not unusual) cases where SCIPCM or IPCM may be computationally expensive or poorly convergent, similar results may be obtained by using PCM to generate the dielectric field, followed by computation of ΣV_s^2 or ΔU_q^I at the 0.0004 or 0.0001 e^-/bohr^3 SAS_I. Unlike previous group contribution approaches,

the model could be practically applied to any moderately small (≤20 non-hydrogen atoms) molecule containing C, H, N, O, S, F, Cl, and Br.

The uncertainty in predicted ΔG_{solv} values propagated from error in S_{pred} calculations depends on the magnitude of the LSER coefficient, *s*, in eq 1. The *s* term indicates the change in electrostatic interaction that the solute will experience in going between the two solvation environments, as defined by ΔG_{solv}. The largest documented *s* value is probably that for air−water partitioning, where $s = 2.55$.[47] In this limiting case one may therefore expect a typical log *P* error of $\approx s\sigma_S = 2.55 \times 0.11 = 0.28$, or a factor of 1.9 in the partition coefficient, as a result of the uncertainty in the *S* model proposed here.

HF/MIDI! and B3LYP/6-311G(2df,2p) molecular orbital computations of gas phase dipole moments compared favorably to measurement data with correlation coefficients of 0.96 and 0.975, respectively. Given this performance for charge distribution estimates of small and medium sized molecules, we do not expect significantly better results for prediction of the electrostatic variable in S_{pred}. In addition to molecular orbital model limitations, the largest source of error in S_{pred} values is probably contamination by solute−solvent hydrogen-bonding interactions inherent in the original development of solute *S* values. This small amount of blending of *S* with other physical processes probably also reflects the extent to which the LSER assumption of linearly separable physical processes is simply inappropriate.

In future work, extension and validation of the model using a wider range of elements (such as Si, P, and I, to which the MIDI! basis set has been extended[52]) and more multifunctional, biological molecules (such as pesticides and pharmaceuticals) would contribute added insight and utility to this investigation. Additionally, development of general LSER approaches which rely on more physically transparent parameters such as computed electrostatics descriptors may offer insights. Such studies could more deeply evaluate the assumptions and limitations of the LSER approximation, thereby leading to a better understanding of these highly successful but preponderantly empirical models.

Acknowledgment. This research was supported by the EPA/ONR/NCERQA Science to Achieve Results Award R-82902301-0, the Alliance for Global Sustainability, the Martin Family Society of Fellows for Sustainability, and the Ralph M. Parsons Fellowship Foundation. The authors also acknowledge Drs. Bernhardt L. Trout, Bruce Tidor, Daniel Blankschtein, Elfatih A. B. Eltahir, and Kenneth Beers for their insightful comments, modeling guidance, and support with molecular computations. Additional thanks goes to Dr. Michael H. Abraham for helpful conversations about this project, to Pharma Algorithms for granting an evaluative license of the AbSolv 2.0 software, and to Dr. Donnan Steele for useful comments on the manuscript.

References and Notes

(1) Kamlet, M. J.; Abboud, J.-L. M.; Abraham, M. H.; Taft, R. W. *J. Org. Chem.* **1983**, *48*, 2877−2887.

(2) Kamlet, M. J.; Doherty, R. M.; Abraham, M. H.; Marcus, Y.; Taft, R. W. *J. Phys. Chem.* **1988**, *92*, 5244−5255.

(3) Abraham, M. H.; Poole, C. F.; Poole, S. K. *J. Chomatogr. A* **1999**, *842*, 79−114.

(4) Goss, K.-U.; Schwarzenbach, R. P. *Environ. Sci. Technol.* **2001**, *35*, 1−9.

(5) Cramer, C. J.; Famini, G. R.; Lowrey, A. H. *Acc. Chem. Res.* **1993**, *26*, 599−605.

(6) Goss., K.-U. *J. Phys. Chem. B* **2003**, *107*, 14025−14029.

(7) Kamlet, M. J.; Abboud, J. L.; Taft, R. W. *J. Am. Chem. Soc.* **1977**, *99*, 6027−6038.

(8) Abraham, M. H.; Whiting, G. S.; Doherty, R. M.; Shuely, W. J. *J. Chromatogr.* **1991**, *587*, 213−228.

(9) Abraham, M. H.; McGowan, J. C. *Chromatogr.* **1987**, *23*, 243−246.

(10) Abraham, M. H.; Whiting, G. S.; Doherty, R. M.; Shuely, W. J. *J. Chem. Soc., Perkin Trans. 2* **1990**, *8*, 1451−1460.

(11) Abraham, M. H.; Ibrahim, A.; Zissimos, A. M. *J. Chomatogr. A* **2004**, *1037*, 29−47.

(12) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Duce, P. P.; Morris, J. J.; Taylor, P. J. *J. Chem. Soc., Perkin Trans. 2* **1989**, *6*, 699−711.

(13) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Morris, J. J.; Taylor, P. J. *J. Chem. Soc., Perkin Trans. 2* **1990**, *4*, 521−529.

(14) Taft, R. W.; Abraham, M. H.; Famini, G. R.; Doherty, R. M.; Abboud, J.-L. M.; Kamlet, M. J. *J. Pharm. Sci.* **1985**, *74*, 807−814.

(15) Abraham, M. H. *J. Phys. Org. Chem.* **1993**, *6*, 660−684.

(16) Abraham, M. H. *J. Chromatogr.* **1993**, *644*, 95−139.

(17) Brinck, T.; Murray, J. S.; Politzer, P. *Mol. Phys.* **1992**, *76*, 609−617.

(18) Murray, J. S.; Politzer, P. *J. Mol. Struct.* **1998**, *425*, 107−114.

(19) Zissimos, A. M.; Abraham, M. H.; Klamt, A.; Eckert, F.; Wood, J. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1320−1331.

(20) Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074−5085.

(21) Zou, J.; Yu, Q.; Shang, Z. *J. Chem. Soc., Perkin Trans. 2* **2001**, *8*, 1439−1443.

(22) Lewis., D. F. V. *J. Comput. Chem.* **1987**, *8*, 1084−1089.

(23) Lamarche, O.; Platts, J. A.; Hersey, A. *Phys. Chem. Chem. Phys.* **2001**, *3*, 2747−2753.

(24) Svozil, D.; Sevcik, J. G. K.; Kvasnicka, V. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 338−342.

(25) Lowrey, A. H.; Cramer, C. J.; Urban, J. J.; Famini, G. R. *Comput. Chem.* **1995**, *19*, 209−215.

(26) Platts, J. A.; Butina, D.; Abraham, M. H.; Hersey, A. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 835−845.

(27) Weckwerth, J. D.; Vitha, M. F.; Carr, P. W. *Fluid Phase Equilib.* **2001**, *183−184*, 143−157.

(28) Wangsness, R. K. *Electromagnetic Fields*; John Wiley & Sons: New York, 1979.

(29) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.

(30) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117−129.

(31) Aqvist, J.; Hansson, T. *J. Phys. Chem.* **1996**, *100*, 9512−9521.

(32) Milischuk, A.; Matyushov, D. V. *J. Phys. Chem. A* **2002**, *106*, 2146−2157.

(33) Milischuk, A.; Matyushov, D. V. *J. Chem. Phys.* **2003**, *118*, 1859−1862.

(34) Matyushov, D. V. *J. Chem. Phys.* **2004**, *120*, 1375−1382.

(35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98, revision a.6. Gaussian, Inc.: Pittsburgh, PA, 1998.

(36) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(37) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281−301.

(38) Green, D. F.; Tidor, B. *J. Phys. Chem. B* **2003**, *107*, 10261−10273.

(39) Cramer, C. J. *Essentials of Computational Chemistry. Theories and Models*; John Wiley & Sons, Ltd.: New York, 2002.

(40) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327−335.

(41) Barone, V.; Cossi, M.; Tomasi, J. *J. Comput. Chem.* **1998**, *19*, 404−417.

(42) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441−451.

(43) Marten, B.; Kim, K.; Cortis, C.; Freisner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775−11788.

(44) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098−16104.

(45) Lide, D. R.; Frederikse, H. P. R. *CRC Handbook of Chemistry and Physics*, 76th ed.; CRC Press: Boca Raton, FL, 1995.

(46) Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. *J. Pharm. Sci.* **1994**, *83*, 1085−1100.

(47) Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A. *J. Chem. Soc., Perkin Trans. 2* **1994**, *8*, 1777−1790.

(48) Weast, R. C.; Astle, M. J. *CRC Handbook of Chemistry and Physics*; CRC Press: Boca Raton, FL, 1981.
(49) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C: The Art of Scientific Computing*; Cambridge University Press: Cambridge, U.K., 1993.
(50) Pharma Algorithms. *AbSolv, v. 2.0*. 2004.
(51) Israelachvili, J. N. *Intermolecular and Surface Forces*, 2nd ed.; Academic Press: New York, 1991.
(52) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1998**, *99*, 192−196.