Theor Chem Acc (1997) 98:85–109

Theoretical
Chemistry Accounts

A universal model for the quantum mechanical calculation of free energies of solvation in non-aqueous solvents

David J. Giesen[1], **Gregory D. Hawkins**[1], **Daniel A. Liotard**[2], **Christopher J. Cramer**[1], **Donald G. Truhlar**[1]

[1] Department of Chemistry and Supercomputer Institute, 207 Pleasant St. SE, Minneapolis, MN 55455-0431, USA
[2] Laboratoire de Physicochimie Théorique, Université de Bordeaux 1, 351 cours de la Liberation, F-33405 Talence Cedex, France

Received: 18 June 1997 / Accepted: 14 July 1997

Abstract. The SM5.4 quantum mechanical solvation model has been extended to calculate free energies of solvation in virtually any organic solvent. Electrostatics and solute-solvent polarization are included self-consistently by the generalized Born equation with class IV charges, and first-solvation-shell effects are modeled in terms of solvent-accessible surface areas that depend on solute geometries and four solvent descriptors. The inclusion of solvent properties in the first-solvation-shell term provides a model that predicts accurate solvation free energies in any solvent for which those properties are known. The model was developed using 1786 experimentally measured solvation free energies for 206 solutes in one or more of 90 solvents. Parameters have been obtained for solutes containing H, C, N, O, F, S, Cl, Br, and I, and the solutes used for parameterization span a wide range of organic functional groups. Solvents used in the parameterization contain H, C, N, O, F, P, S, Cl, Br, and I and include the most common organic solvents. Two general parameterizations are presented here, one for use with the AM1 Hamiltonian (SM5.4/AM1) and one for use with the PM3 Hamiltonian (SM5.4/PM3). In each case, one parameter is specially re-optimized for benzene and toluene to reduce systematic errors for these solvents. Chloroform is also treated with special parameters. The final mean unsigned error for both the SM5.4/AM1 and SM5.4/PM3 parameterizations is less than 0.5 kcal mol^{-1} over the entire data set of 1786 free energies of solvation in 90 organic solvents.

Key words: Solvation model – Solvents, organic – Electrostatic effects – Partial charges, class IV – Surface tensions, atomic

1 Introduction

The interaction of a solute molecule with the surrounding solvent or the transfer of a solute from one solvent to another can dramatically change the solute properties, including free energy, reaction rates, reaction paths, and even the identity of the solute molecule itself [1]. Thus the modeling of solute-solvent interactions in a range of solvents is a critical area for the development of predictive techniques in theoretical chemistry. While much work has been done by many groups in developing quantum mechanical models for water, as reviewed elsewhere [2–6], much less effort has been devoted to developing such models for non-aqueous solvents [7–14]. For some non-aqueous solvents, this is due to a lack of experimental data, although a large body of data is available for 1-octanol [15] and *n*-hexadecane [16, 17], and a significant amount is also available for other alkanes, cyclohexane, benzene, toluene, xylenes, ethyl ether, chloroform, carbon tetrachloride, and chlorobenzene [18]. The sparsity of data for other common solvents, however, is a serious impediment to solvent-by-solvent approaches for developing and validating solvation models for most individual solvents. But, if one groups together data from *all* organic solvents, then a very large number of data are available [18], and in this paper we show that it is possible to analyze these data as a whole and develop a model that encompasses a large number of solvents in a single framework. This approach eliminates the need for extensive amounts of data measured in one solvent, and it allows us to predict solvation energies in solvents for which little or no experimental data exist (which is often the case for solvents miscible with water).

Solvation energies for non-aqueous solvents are important for modeling a wide variety of phenomena in organic chemistry. In addition, in conjunction with an aqueous solvation model, such as the SM5.4/AM1 or SM5.4/PM3-aqueous parameterizations we have presented previously [19], they can be used to predict partition coefficients of solutes between an organic phase and water [20, 21], and such partition coefficients are often used to provide some indication of how likely it is for the solute to penetrate a lipid bilayer, skin, brain, central nervous system, or other biophase, or to bind to a non-polar site in or on a protein [22].

The present model is an extension of the SM5.4-aqueous model [19] to organic solvents, and it uses ex-

---

Correspondence to: C.J. Cramer or D.G. Truhlar tended versions of the functional forms developed in that work. A preliminary communication of the present extension has been presented elsewhere [23], and the results for chloroform have also been published separately [24]. (The latter event could be confusing since the chronology of publication does not match the chronology of doing the work; in particular, some of that work builds on work presented in detail in the present paper.) In all cases, we have obtained parameters suitable for use with both the AM1 [25–27] and PM3 [28] electronic Hamiltonians; these parameterizations of the model are differentiated by the notations /AM1 and /PM3, respectively. The work presented in prior publications [19, 23, 24], and here, taken together, constitute the SM5.4/AM1 and SM5.4/PM3 parameterizations of the SM5.4 solvation model for water and organic solvents.

2 Experimental data

For this work, our main source of experimental data was the large database of partition coefficients developed by Leo [18]. The partition coefficient ($P_{\mathrm{A/B}}$) of a solute between solvent A and solvent B is defined as the ratio of the concentration of the solute in A to the concentration of the solute in B when the concentrations are in equilibrium with each other. This is expressed using equilibrium concentrations as

$$P_{\mathrm{A/B}} = \frac{[\mathrm{solute}]_{\mathrm{A}}}{[\mathrm{solute}]_{\mathrm{B}}} \tag{1}$$

The base-10 logarithm of the partition coefficient is related to the standard-state free energy of transfer from solvent B to solvent A by

$$\log P_{\mathrm{A/B}} = -\Delta G^{\circ}_{\mathrm{A/B}}/2.303\,RT \tag{2}$$

where

$$\Delta G^{\circ}_{\mathrm{A/B}} = \Delta G^{\circ}_{\mathrm{A}} - \Delta G^{\circ}_{\mathrm{B}} \ , \tag{3}$$

R is the ideal gas constant, T is the temperature, and ΔG°_{X} is the standard-state solvation free energy of the solute in solvent X. From Eqs. (1)–(3) we can see that for a given solute, if the partition coefficient and the free energy of solvation in one solvent are known, we can calculate the free energy of solvation in the second solvent. For solvent/air partition coefficients, $\Delta G^{\circ}_{\mathrm{B}}$ is intrinsically 0. The ubiquitous "$\log P$" of the rational drug design literature is a special case of Eq. (2) corresponding to A = 1-octanol and B = water; it is often denoted $\log P_{\mathrm{o/w}}$.

Some experimental free energies of solvation for non-aqueous solvents are directly available, in particular for the solvents 1-octanol [29] and *n*-hexadecane [10], but most of the ΔG°_{X} values we used for non-aqueous solvents were extracted from organic-solvent/water partition coefficients. Knowing both the organic-solvent/water partition coefficient and the experimental free energy of solvation of a solute in water, we were able to use Eqs. (1)–(3) to calculate the free energy of solvation in the non-aqueous solvent. For all solute/solvent combinations in this paper, solvation free energies were obtained from one of four sources with the following priorities: (1) free energy of solvation data gathered previously for hexadecane solvent, as described previously [10]; (2) 46 air/1-octanol transfer free energies taken from Dallas' thesis [29]; (3) solvent/air partition coefficients taken from the Medchem database [18]; or (4) solvent/water partition coefficients taken from the Medchem database combined with the water/air data [19] used in the development of the SM5.4-aqueous models. The Medchem database contains many partition coefficients that were measured under conditions that do not correspond to the standard conditions for which this model was developed. Data were rejected for any of the following reasons: (1) measured outside the range of 20–30°C; (2) measured outside the pH range of 6–8; (3) measured in an aqueous phase that was not pure aqueous (with the exception of dihydrogen phosphate buffers); (4) indications that the solute was not in its true form, e.g. dimerization; (5) measurements where salting out occurred; or (6) measurements that were marked unreliable in the Medchem database. For solute/solvent combinations where more than one acceptable partition coefficient existed at the highest available priority level, data more than two standard deviations from the mean were iteratively removed until no more outliers remained, and the mean of the remaining data was used to represent the experimental situation. An examination of cases where there are multiple measurements led to the conclusion that the typical experimental uncertainty is 0.2 kcal/mol.

Data from the following types of solvents were not used: (1) non-homogeneous solvents such as micelles, (2) mixtures of two or more substances, (3) solvents for which less than 5 data existed. This led to 101 solvents that were used in preliminary work. This list was further reduced to 90 solvents by the requirement that all five solute properties selected for parameterization (see Sect. 4) were either known or easily estimable. These 90 solvents comprise the organic SM5.4 non-aqueous solvent training set.

The set of solutes used in the present paper was selected from the previously described [19] "meta set," which consists of about 240 molecules considered for use in parameterizing the aqueous [19] model. This set consists of organic and small inorganic molecules which contain H, C, N, O, F, S, Cl, Br, and/or I atoms and spans a very wide variety of common organic functional groups. After deleting molecules with multiple or ambiguous conformations, as described previously [19], we were able to extract free energies of solvation for 206 of the remaining molecules in one or more solvents of the SM5.4 non-aqueous solvent training set, and the total number of non-aqueous free energies of solvation used in developing the final model was 1786. One solute, hydrazine (N_2H_4), was only used in the development of the organic SM5.4/AM1 model; it was removed from the organic SM5.4/PM3 development set because the PM3 minimum-energy structure has a qualitatively different H—N—N—H torsion angle from the experimental one [30]. Since there were data for hydrazine in four solvents, the number of free energies of solvation used to develop the organic SM5.4/PM3 model was thus 1782.

3 Background

The model presented here is a member of the SMx ($x = 1, 2, \ldots$) solvation model family and will be called the SM5.4 model for reasons delineated below. Like all SMx models [9–11, 19, 23, 24, 31–38], it is based on treating the solvent as a continuum dielectric. The solvent molecules, including those in the first solvation shell (the solute boundary for which is determined by a set of solute atomic radii called coulomb radii), are replaced by a continuous dielectric medium with a dielectric constant equal to that measured for the bulk solvent. The solute is allowed to relax both its nuclear geometry and its electronic wave function in the presence of the solvent, and first-solvation-shell effects beyond those already included through the portion of the dielectric medium in the first solvation shell are included via empirical atomic surface tensions. These first-solvation-shell effects include the deviation of the dielectric constant from its bulk value in the region near the solute as well as short-range solute-solvent interactions. The first-solvation-shell terms also make up for non-uniqueness of the coulomb radii and indeed for the deficiencies of modeling the solute electron density with nuclear-centered monopoles and of modeling the solute-solvent boundary as sharp. The nomenclature of the parameterizations (SM5.4/AM1 and SM5.4/PM3) refers to the geometry-dependent functional forms for atomic surface tensions (SM5), the method of calculating partial charges (.4 denotes class IV methods for atomic partial charges [39]), and the underlying Hamiltonian (/AM1 and AM1 and /PM3 for PM3) for which the parameters were developed. We note that these SM5.4 parameterizations are a subset of a larger number of parameterizations that use the basic SM5 framework for atomic surface tensions; the full set of such parameterizations constitutes the SM5 family of models. Further SM5 parameterizations that use the extensions of the strategies explained in this paper for treating organic solvents in a universal way are under development already, for example, versions parameterized for use with the MNDO/d method [40–42] or density functional theory [43, 44].

The models presented in this paper draw extensively on previous models we have developed [10, 11, 19, 45], and so the reader will at times be directed to previous papers for computational details or parameters. We begin by writing the free energy of solvation as

$$\Delta G_{\mathrm{S}}^{\circ} = \Delta G_{\mathrm{ENP}} + G_{\mathrm{CDS}} \quad . \tag{4}$$

The first term, ΔG_{ENP}, is the sum of free energies associated with (1) the change from the gas phase in the solute energy associated with altering the solute electronic and nuclear wave functions, (2) the change in the solute atomic electrostatic self energies and charge-charge interactions due to solvent dielectric field screening of the solute charges, and (3) the reorganization cost of the solvent. We assume that any change in the rotational or vibrational internal free energy of the solute upon solution is negligible, although in principle it can be included in ΔG_{ENP}. G_{CDS} is the free energy associated with short-range solute-solvent and solvent-solvent intermolecular interactions such as dispersion, cavitation, and solvent structural rearrangement. The non-electrostatic components of hydrogen bonding effects are also included in G_{CDS}. All calculations in this work refer to a standard state ideal concentration of $1\,\mathrm{mol\,l^{-1}}$ in both the gas and liquid solution phase and a temperature of 298 K; if we used different concentrations for the gas-phase and solution standard states, we would need another term in Eq. (4) to account for their difference.

As explained above, the first term in Eq. (4), ΔG_{ENP}, is a sum of contributions. Computationally, we write

$$\Delta G_{\mathrm{ENP}} = \Delta E_{\mathrm{EN}} + G_{\mathrm{P}} \tag{5}$$

where ΔE_{EN} is the change in the electronic and nuclear energy of the solute upon relaxation from the gas-phase minimum energy structure to the solution-phase minimum energy structure, and G_{P} is the polarization free energy due to the screening of the solute charges by the solvent dielectric field, the solute-solvent electrostatic interactions, and the solvent-solvent dielectric polarization cost.

The calculation of ΔE_{EN} derives from solvent-induced changes in molecular geometry and electronic structure according to

$$\begin{aligned}\Delta E_{\mathrm{EN}} = & \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}^{(\mathrm{s})}\left(H_{\mu\nu}^{(\mathrm{s})} + F_{\mu\nu}^{(\mathrm{s})}\right) + \frac{1}{2}\sum_{k,k'\neq k} \frac{Z_k Z_{k'}}{r_{kk'}^{(\mathrm{s})}} \\ & - \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}^{(g)}\left(H_{\mu\nu}^{(g)} + F_{\mu\nu}^{(g)}\right) - \frac{1}{2}\sum_{k,k'\neq k} \frac{Z_k Z_{k'}}{r_{kk'}^{(g)}}\end{aligned} \tag{6}$$

where **P**, **H**, and **F** are, respectively, the density, one-electron, and Fock matrices, μ and ν run over valence atomic orbitals, Z_k is the valence nuclear charge of atom k, and $r_{kk'}$ is the internuclear separation between atoms k and k'. The superscripts (s) and (g) distinguish values in solution from those in the gas phase. In the cases of **H** and $r_{kk'}$, gas-phase and solution values differ only if the geometry in solution is different from that in the gas phase. However, in *addition* to any difference due to geometry change, the density matrix **P** will be different in solution because it is determined from molecular orbitals derived using the self-consistent reaction field approach, which minimizes ΔG_{ENP} as discussed further below. Since the Fock matrix **F** contains two-electron terms that depend on the density matrix, it too can show differences between the gas phase and solution even in the absence of geometric differences.

For a given model of the boundary between the solute and the dielectric medium, G_{P} can be calculated by using the Poisson equation [46–51]; however, we approximate it using the generalized Born equation [35, 45, 52–58].

$$\begin{aligned}G_{\mathrm{P}} = & -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_{k,k'} q_k q_{k'} \\ & \left(r_{kk'}^2 + \alpha_k \alpha_{k'} \exp\left[-r_{kk'}^2 / d_{kk'} \alpha_k \alpha_{k'}\right]\right)^{-\frac{1}{2}} \quad .\end{aligned} \tag{7}$$

In Eq. (7), ε is the bulk dielectric constant of the solvent, q_k is the partial atomic charge on atom k, $r_{kk'}$ is the distance between atoms k and k', α_k is the effective coulomb radius of atom k, and $d_{kk'}$ is an empirically optimized constant whose value depends on the atomic pairs k and k'. The effective coulomb radius of atom k is calculated from the intrinsic coulomb radius of atom k according to the dielectric screening approximation of Still et al. [57], which accounts for the fact that surrounding portions of the solute prevent the solvent from fully screening the electrostatic interactions of atom k. The intrinsic coulomb radii are parameters originally developed primarily to obtain good aqueous solvation energies for ions in water [19], and these parameters, along with the set of $d_{kk'}$, also developed in the aqueous model, are retained in the present extension to organic solvents, both in functional form and value, without re-optimization from the values used in the aqueous parameterizations. The intrinsic coulomb radii are constants. In earlier versions of the SMx models, some coulomb radii depended on partial atomic charge. This dependence was ultimately deemed unnecessary, and it is moreover undesirable, since it makes derivatives of Eq. (7) with respect to changes in the density matrix (required for the self-consistent reaction field method, vide infra) more complicated.

The partial atomic charges are obtained using the class IV CM1 mapping of NDDO Mulliken charges [10, 39]. The specific mapping used is determined by the underlying Hamiltonian; CM1A [39] is used with AM1 charges for the SM5.4/AM1 model, and CM1P [39] is used with PM3 charges for the SM5.4/PM3 model. The CM1 mappings were designed to yield charges that better reproduce dipole moments than do the NDDO Mulliken charges at roughly the same cost as the zero-differential-overlap Mulliken population analysis. The use of CM1 charges allows a more physical splitting between the ΔG_{ENP} term and the G_{CDS} term than that given by conventional population analysis; this will be important for developing a physical model for calculating G_{CDS} using solvent properties and for using the dielectric constant to predict the solvent dependence of G_{P}.

The ΔG_{ENP} term is optimized self-consistently by the self-consistent reaction-field method [55]. By including in the Fock matrix the derivative of Eq. (7) with respect to changes in the density matrix, we are able to self-consistently optimize the solute geometry and electronic wave function in the presence of the solvent reaction field, which is the electrostatic field at the solute due to dielectric polarization of the solvent. The calculation of CM1 charges [39] and their incorporation into the NDDO Fock matrix to allow self-consistent polarization of the solute electronic wave function [10] are detailed elsewhere.

The second term in Eq. (4) G_{CDS}, involves atomic and molecular surface tensions. The SM5 approach to atomic surface tensions was first developed for aqueous solutions [19], and in the organic solvation model here it is extended in two ways: (1) we add intermediate-range first-solvation-shell effects as in the SM4 model for alkane solvents [10, 11], and (2) we allow the surface tension coefficients to depend on solvent properties. In particular, we write

$$G_{\mathrm{CDS}} = G_{\mathrm{CD}} + G_{\mathrm{CS}} \tag{8}$$

$$G_{\mathrm{CD}} = \sum_{k} \sigma_k A_k\left(\{R_Z\}, R_{\mathrm{S}}^{\mathrm{CD}}\right) \tag{9}$$

and

$$G_{\mathrm{CS}} = \sigma^{\mathrm{CS}} \sum_{k} A_k\left(\{R_Z\}, R_{\mathrm{S}}^{\mathrm{CS}}\right) \tag{10}$$

where σ_k is the geometry-dependent cavity-dispersion (CD) atomic surface tension of atom k, $A_k\left(\{R_Z\}, R_{\mathrm{S}}^{\mathrm{CD}}\right)$ is the solvent-accessible surface area [59] of atom k calculated with the set $\{R_Z\}$ of solute atomic radii and with effective solvent radius $R_{\mathrm{S}}^{\mathrm{CD}}$, σ^{CS} is the cavity-structural (CS) surface tension of the whole solute, and $A_k\left(\{R_Z\}, R_{\mathrm{S}}^{\mathrm{CS}}\right)$ is the solvent-accessible surface area of atom k calculated with the same set of solute atomic radii and with effective solvent radius $R_{\mathrm{S}}^{\mathrm{CS}}$. All SM5.4 models use the same values of the van der Waals radii as used in the SM5.4-aqueous models [19], namely the values of Bondi [60]. The solvent effective radii and surface tensions will be explained below, and the details of the algorithm for calculating solvent-accessible surface areas are given in a previous paper [35].

The specific form of the atomic surface tensions is determined by the solvation model. In particular, in SM1, they depend only on atomic number; in SM1a, they depend on manually assigned atomic type (as in molecular mechanics "types," e.g., carbonyl O or alcohol O, etc.); in SM2–SM3.1 and SM4, they depend on bond orders; and in the SM5.4-aqueous models, they depend on bond distances. In contrast, σ^{CS} is independent of solute properties. The SM5.4 model for organic solvents uses precisely the same bond-distance dependencies as the SM5.4-aqueous models, and all σ_k and σ^{CS} depend on solvent properties too.

In SM5.4 models, the geometry dependencies in σ_k are designed to recognize nearest-neighbor (and, in the case of N, next-nearest-neighbor) bonding patterns. The overall surface tension for atom i is calculated as

$$\sigma_k = \tilde{\sigma}_{Z_k} + \sum_{k'} \left[\tilde{\sigma}_{Z_k Z_{k'}} f_{Z_k Z_{k'}}(r_{kk'}) + \delta_{Z_k 6} \delta_{Z_{k'} 6} \tilde{\sigma}^{(2)}_{Z_k Z_{k'}} f^{(2)}_{Z_k Z_{k'}}(r_{kk'})\right] \tag{11}$$

where $\tilde{\sigma}_{Z_k}$, $\tilde{\sigma}_{Z_k Z_{k'}}$, and $\tilde{\sigma}^{(2)}_{Z_k Z_{k'}}$ are composite surface tension coefficients associated with atoms with atomic numbers Z_k and $Z_{k'}$, and $f_{Z_k Z_{k'}}$ is a function of interatomic distances $r_{kk'}$. Table 1 lists the bonding partners recognized by the SM5.4 geometry dependencies and the most common functional groups that are affected by each type. The bonding functions are zero over a semi-infinite range of distance from some cutoff to infinity, but they become non-zero when $r_{kk'}$ corresponds to a value slightly larger (e.g., 0.3 Å) than its typical value in the target functional group(s). For example, f_{18} is non-zero for all hydrogens that are within a few tenths of an Å of a typical bonding distance of an oxygen. Note that $f^{(2)}_{ZZ'}(r_{kk'})$ is non-zero only for alkyne carbons. The interatomic distance functions and their parameters are the same for the present models as for the earlier SM5.4-aqueous models [19].

Table 1. Atomic bonding patterns recognized by the SM5 formalism

Z—Z′	Primarily affected functional groups
H—C	Hydrogen bound to carbon
H—N	Hydrogen bound to nitrogen
H—O	Hydrogen bound to oxygen
H—S	Hydrogen bound to sulfur
C—C	Any carbon bound to carbon
C—C$^{(2)}$	Alkyne carbon
O—C	Carbonyl oxygen
O—O	Oxygen geminal to oxygen (carboxylic acids, esters)
N—C	Tertiary amine > secondary amine > aromatic amine > primary amine
O—N	Oxygen bound to nitrogen
S—S	Sulfur bound to sulfur

In the SM5.4-aqueous model the second term of Eq. (8) is not present, and the $\tilde{\sigma}_Z$, $\tilde{\sigma}_{ZZ'}$, and $\tilde{\sigma}^{(2)}_{ZZ'}$ coefficients are primary fitting parameters. In the extension of the SM5.4 model to organic solvents, $\tilde{\sigma}_Z$, $\tilde{\sigma}_{ZZ'}$, $\tilde{\sigma}^{(2)}_{ZZ'}$, and σ^{CS} are derived parameters obtained by expressing them as bilinear functions of primary parameters and solvent descriptors. This aspect of the present model is explained in the next section.

4 Model development

4.1 Solvent descriptors

The original SM5 formalism was extended in two ways to make a general model for non-aqueous solvents. The first extension was to include an intermediate-range molecular surface tension term similar to one used successfully in our previous alkane models [10, 11]. The second extension was to make the surface tension coefficients be functions of solvent properties, thereby allowing the model to account for short- and intermediate-range solute-solvent interactions that depend on solvent properties.

In organic solvents, $R_{\mathrm{S}}^{\mathrm{CD}}$ is taken to be [19] 1.7 Å, and $R_{\mathrm{S}}^{\mathrm{CS}}$ is taken to be 3.4 Å. The intermediate-range solute-solvent interaction that we identified in our parametric model appears to be reasonably independent of the atomic constitution of the surface of the solute. This interaction may be a model-dependent effect arising from the way we put the model together, or it may be physical. Taking the latter point of view, we assume that it arises from the structural re-organization energy required to form a cavity in the solvent, and its large, positive value is consistent with this picture [10]. For reasons that are unclear, but that are probably associated with the small size of the water molecule compared to the sizes of organic solvents, such a term was not needed in the development of the aqueous models. We rationalize this by thinking of $R_{\mathrm{S}}^{\mathrm{CS}}$ as equal to $R_{\mathrm{S}}^{\mathrm{CD}}$ for water, which is a uniquely small solvent molecule. As a consequence, in water, the σ^{CS} terms are effectively lumped into the $\tilde{\sigma}_Z$ CD terms, resulting in net $\tilde{\sigma}_Z$ parameters for the resultant combined CD and CS effects.

It is reasonable that $R_{\mathrm{S}}^{\mathrm{CD}}$ is the same for water and for organic solvents if we think of it as representing the range of dispersion interactions. We note that the final free energies predicted by the model are not overly sensitive to increasing the solvent radii. For example, if we used 2.0 Å and 4.0 Å instead of 1.7 Å and 3.4 Å, all surface tension coefficients would be somewhat smaller, but the predicted free energies of solvation would not be very much different.

The critical extension required to take the single solvent formalism to a multiple solvent formalism is to make the surface tension coefficients be functions of the properties of the solvent. The most important decision in the development of the organic SM5.4 models was the choice of solvent properties to use for parameterizing the atomic surface tensions in Eqs. (9)–(11). The solvent properties should be chosen with the following criteria in mind: (1) they should describe the full range of important solute-solvent interactions, (2) they should be readily available for a wide variety of solvents, and (3) it should be easy to interpret the solute-solvent interactions they represent. The goal of this stage of the work was to identify parameters P_i for atomic surface tensions σ_α of atom type α to depend on, as in

$$\sigma_\alpha = \sum_{i=1}^{3\text{ or }4} \sigma_{\alpha i} P_i \tag{12}$$

A large body of work [1, 61–73] has led to the identification of various sets of two to seven seemingly independent solvent characteristics, but four such non-linearly-dependent solvent parameters seem to be needed in general to capture the most important differences between solvents. The two most important solvent characteristics appear to be "polarity" and "acidity," either Lewis acidity (electron-pair accepting ability) or H-bond donating ability. These characteristics are typically quantified by some function of dielectric constant or by $\mathrm{E_T}(30)$, where the latter is the transition energy of Reichardt's dye no. 30 [1]. It is generally agreed that $\mathrm{E_T}(30)$ measures some mixture of "polarity" and "acidity" (with a little bit of "polarizability"– see below – thrown in). A third obviously important solvent property is basicity, nucleophilicity, H-bond accepting ability, or electron-pair donating ability. Recommended possibilities (none fully ideal) include the ionization potential or one of the many available solvent descriptors that correlates qualitatively with chemical basicity (although most of the successful basicity scales, e.g., the DN donor number scale, appear to be unavailable for many solvents of interest). A fourth important solvent characteristic that has emerged from previous work is polarizability, typically measured by some function of the index of refraction for sodium D line radiation (n_D). A fifth candidate property is macroscopic surface tension γ_{m} of the solvent; in previous work [11], we have shown that experimental trends in the free energy of solvation for alkane solvents can be successfully reproduced using a linear function of γ_{m}. Based on this background, we initially looked at a variety of solvent descriptors, including $\mathrm{E_T}(30)$ [1], donor number and acceptor number (DN and AN) [1, 73], Abraham's $\sum \alpha_2^{\mathrm{H}}$

and $\sum \beta_2^{\mathrm{H}}$ scales [74], dielectric constant (ε) [75], n_D [75], various functions of dielectric constant and index of refraction (such as $\frac{n_D^2-1}{n_D^2+2}$, $\frac{n_D^2-1}{2n_D^2+1}$, $1-\frac{1}{\varepsilon}$, $\frac{\varepsilon-1}{\varepsilon+2}$, $\frac{\varepsilon-1}{2\varepsilon+1}\frac{n_D^2+1}{2n_D^2+1}$, etc.), polarizability divided by molecular volume [75], γ_{m} [75], internal pressure [76], cohesive energy density [76], Hildebrand's solubility parameter [76], and boiling point. We examined fits with anywhere from zero to four solvent properties and tried numerous functions of the solvent properties and several cross terms involving two or more properties. We found that it is most important to have descriptors for solvent acidity, basicity, and polarizability. Apparently because solvent polarity is already accounted for through the appearance of $\left(1-\frac{1}{\varepsilon}\right)$ in ΔG_{ENP}, including ε as a solvent property in the calculation of G_{CDS} did not significantly improve the fit. The fact that including ε in the fit did not offer a significant improvement is encouraging since such improvement might have indicated a non-physical partitioning between ΔG_{ENP} and G_{CDS}.

We found that several different sets of acidity, basicity, and polarizability parameters offered fits of similar quality. We finally chose linear dependencies of the short-range surface tensions on n_D and on Abraham's $\sum \alpha_2^{\mathrm{H}}$ and $\sum \beta_2^{\mathrm{H}}$. The fact that solvation free energy depends linearly on $\sum \alpha_2^{\mathrm{H}}$ and $\sum \beta_2^{\mathrm{H}}$ is not surprising since these parameters were originally developed as components of linear free energy relations ($\sum \beta_2^{\mathrm{H}}$ was developed primarily from iterative regressive analysis on water-solvent partition coefficients). However, a linear dependence of solvation free energies on n_D is harder to justify theoretically. The index of refraction is related to the polarizability [75], and therefore it can be used as a measure of the ability of the solvent to undergo dispersion interactions with the solute. There are several functions, typically quadratic or rational, which relate n_D to polarizability, but we found that linear dependence on n_D performed as well as or better than linear dependence on any of these functions.

The hydrogen bond donating ability (acidity) of the solvent molecule is described by $\sum \alpha_2^{\mathrm{H}}$, and the hydrogen bond accepting ability (basicity) of the solvent molecule is described by $\sum \beta_2^{\mathrm{H}}$. Although these scales of acidity and basicity were developed for solute molecules [74], we found them to give useful results as a relative scale for the same molecules as solvents, even for polyfunctional solvents. The three descriptors discussed here are widely available for the solvents we examined, and no preliminary fit performed better than the combination of n_D, $\sum \alpha_2^{\mathrm{H}}$, and $\sum \beta_2^{\mathrm{H}}$. A major advantage of using Abraham's parameters is that they are available for most solvents, and methods exist to measure [74, 77–79] or estimate [80] them when they are not available.

The formation of a cavity involves the loss of favourable solvent-solvent dispersion interactions, and dispersion interactions are related to the polarizability of the molecules involved, so we also added a dependence of σ^{CS} on n_D. In addition, the macroscopic surface tension γ_{m} represents the energy to make a macroscopic cavity in the solvent, and so σ^{CS} was also made to depend on γ_{m}. We found that dependence of σ^{CS} on the macroscopic surface tension improved the fit for several classes of solvents, particularly the solvents which contain nitro groups. Making σ^{CS} additionally dependent on α and β did not improve the model, and so such dependence was not included.

As a consequence of all these considerations we used the following functional forms:

$$\tilde{\sigma}_Z = \sum_{\lambda=n,\alpha,\beta} \hat{\sigma}_Z^{(\lambda)} \lambda \tag{13}$$

$$\tilde{\sigma}_{ZZ'} = \sum_{\lambda=n,\alpha,\beta} \hat{\sigma}_{ZZ'}^{(\lambda)} \lambda \tag{14}$$

$$\tilde{\sigma}_{ZZ'}^{(2)} = \hat{\sigma}_{ZZ'}^{(2,n)} n \tag{15}$$

and

$$\sigma^{\mathrm{CS}} = \sum_{\lambda=n,\gamma} \hat{\sigma}^{\mathrm{CS},\lambda} \lambda \tag{16}$$

where $\hat{\sigma}_Z^{(\lambda)}$, $\hat{\sigma}_{ZZ'}^{(\lambda)}$, $\hat{\sigma}_{ZZ'}^{(2,n)}$, and $\hat{\sigma}^{\mathrm{CS},\lambda}$ are the ultimate surface tension coefficients, and where

$$n = n_D,\ \alpha = \sum \alpha_2^{\mathrm{H}},\ \beta = \sum \beta_2^{\mathrm{H}},\ \gamma = \gamma_{\mathrm{m}}/\gamma_{\mathrm{o}}\ , \tag{17}$$

and γ_{o} is taken as $1\,\mathrm{cal\,mol^{-1}\,\AA^{-2}}$. Note that all four descriptors are unitless.

The solvent properties used in organic SM5.4 models are all widely available properties [74, 75]. In a few cases, the values used here had to be estimated, and this was done either by fits to properties of similar compounds or by extrapolation from data for the solvent of interest at temperatures other than 25°C. The values of the descriptors used for all 90 solvents comprising the organic SM5.4-non-aqueous solvent training set are given in Table 2.

To effect the parameterization, we first determined $\hat{\sigma}_{\mathrm{H}}^{n}$ from the experimental free energy of solvation of the H_2 molecule in n-hexadecane and 1-octanol. The errors in this step are less than $0.1\,\mathrm{kcal\,mol^{-1}}$. Then this coefficient was frozen and all the remaining surface tension coefficients were found by a simultaneous fit to the remaining 1784 data for AM1 or the remaining 1780 data for PM3. This is accomplished by writing

$$G_{(\mathrm{Residual})} = \Delta G^{\circ}_{\mathrm{S(experimental)}} - \Delta G_{\mathrm{ENP}}\ . \tag{18}$$

The set of $\tilde{\sigma}$ are then found by a least squares multiple linear regression of G_{CDS} against $G_{(\mathrm{Residual})}$.

Not all possible coefficients allowed in Eqs. (14) and (15) were included in the regression; the omitted ones were assigned a value of zero. All exposed atoms undergo dispersion interactions with the solvent to at least some extent, and the formation of a cavity involves loss of solvent-solvent dispersion interactions, so all $\hat{\sigma}_Z^{(n)}$ and $\hat{\sigma}_{ZZ'}^{(n)}$ coefficients (for recognized ZZ' pairs, as listed in Table 1) were allowed to be non-zero. In contrast, the choice for which surface tensions should depend on α and β is less obvious. For instance, the surface tension of a hydrogen bonded to an oxygen or nitrogen should depend on the basicity of the solvent (β) but probably not on the acidity (α). The alkyne term did not show any significant dependence on $\sum \alpha_2^{\mathrm{H}}$ and $\sum \beta_2^{\mathrm{H}}$ so this term was made a function of n_D only. Through a process of

Table 2. Solvent descriptors for the SM5.4-non-aqueous solvent training set[a,b]

Solvent	n	α	β	γ	ε
1,2,4-Trimethylbenzene	1.5048	0.00	0.19	42.0	2.37
1,2-Dibromoethane	1.5420	0.10	0.10	56.9	4.93
1,2-Dichloroethane	1.4448	0.10	0.11	45.9	10.19
1,3,5-Trimethylbenzene[c]	1.4994	0.00	0.19	39.7	2.27
1-Bromooctane	1.4524	0.00	0.12	41.3	5.02
1-Butanol	1.3993	0.37	0.48	35.9	17.33
1-Chlorohexane	1.4199	0.00	0.10	37.0	5.95
1-Decanol	1.4372	0.37	0.48	41.0	7.53
1-Fluoroctane	1.3935	0.00	0.10	33.9	3.89
1-Heptanol	1.4249	0.37	0.48	38.2	11.32
1-Hexanol	1.4178	0.37	0.48	37.1	12.51
1-Iodohexadecane	1.4806	0.00	0.15	46.5	3.53
1-Nonanol	1.4333	0.37	0.48	40.1	8.60
1-Octanol	1.4295	0.37	0.48	39.0	9.87
1-Pentanol	1.4101	0.37	0.48	36.5	15.13
1-Propanol	1.3850	0.37	0.48	33.6	20.52
2,2,4-Trimethylpentane	1.4513	0.00	0.00	26.4	1.94
2,6-Dimethylpyridine	1.4953	0.00	0.63	44.6	7.17
2-Butanol	1.3978	0.33	0.56	32.4	15.94
2-Methoxyethanol	1.4024	0.30	0.84	44.4	17.11
2-Methylpyridine	1.4957	0.00	0.58	47.5	9.95
4-Methyl-2-pentanone	1.3962	0.00	0.51	33.3	12.89
Acetic acid	1.3720	0.61	0.44	39.0	6.25
Acetonitrile	1.3442	0.07	0.32	41.3	37.50
Acetophenone	1.5372	0.00	0.48	56.2	17.44
Aniline	1.5863	0.26	0.41	60.6	6.89
Anisole	1.5174	0.00	0.29	50.5	4.22
Benzene	1.5011	0.00	0.14	40.6	2.27
Benzonitrile	1.5289	0.00	0.33	44.3	25.59
Benzyl alcohol	1.5396	0.33	0.56	50.1	11.92
Bromobenzene	1.5597	0.00	0.09	50.7	5.40
Bromoethane	1.4239	0.00	0.12	34.0	9.02
Bromoform	1.6005	0.15	0.06	64.6	4.25
Butanone	1.3788	0.00	0.51	34.5	18.25
Butyl acetate	1.3941	0.00	0.45	35.7	4.99
Butyl ether	1.3992	0.00	0.45	32.3	3.05
Butylbenzene	1.4898	0.00	0.15	41.3	2.36
Carbon disulfide	1.6319	0.00	0.07	45.5	2.61
Carbon tetrachloride	1.4601	0.00	0.00	38.0	2.23
Chlorobenzene	1.5241	0.00	0.07	47.5	5.70
Cyclohexane	1.4266	0.00	0.00	35.5	2.02
Cyclohexanone	1.4507	0.00	0.56	49.8	15.62
Decalin	1.4753	0.00	0.00	43.9	2.20
Dimethylsufoxide	1.4170	0.00	0.88	61.8	46.83
Ethanol	1.3611	0.37	0.48	31.6	24.85
Ethoxybenzene	1.4959	0.00	0.32	46.3	4.18
Ethyl acetate	1.3723	0.00	0.45	33.7	5.99
Ethyl ether	1.3526	0.00	0.41	24.0	4.24
Ethylbenzene	1.4959	0.00	0.15	41.4	2.43
Fluorobenzene	1.4684	0.00	0.10	38.4	5.47
Iodobenzene	1.6200	0.00	0.12	55.7	4.55
Isobutanol	1.3955	0.37	0.48	32.4	16.78
Isopropanol	1.3776	0.33	0.56	30.1	19.26
Isopropyl ether	1.3679	0.00	0.41	24.9	3.81
Isopropylbenzene	1.4915	0.00	0.16	39.8	2.37
m-Cresol	1.5438	0.57	0.34	51.4	12.70
Methylene chloride	1.4242	0.10	0.05	39.1	8.82
N,*N*′-Dimethylacetamide	1.4380	0.00	0.78	47.7	37.78
N,*N*′-Dimethylformamide	1.4305	0.00	0.74	50.7	37.22
n-Decane	1.4102	0.00	0.00	33.6	1.99
n-Dodecane	1.4216	0.00	0.00	35.9	2.01
n-Heptane	1.3878	0.00	0.00	28.3	1.91
n-Hexadecane	1.4345	0.00	0.00	38.9	2.06
n-Hexane	1.3749	0.00	0.00	25.7	1.88
N-Methylformamide	1.4319	0.40	0.55	55.7	181.56
n-Nonane	1.4054	0.00	0.00	32.2	1.96
n-Octane	1.3974	0.00	0.00	30.4	1.94
n-Pentadecane	1.4315	0.00	0.00	38.3	2.03
n-Pentane	1.3575	0.00	0.00	22.3	1.84
n-Undecane	1.4398	0.00	0.00	34.8	1.99
Nitrobenzene	1.5562	0.00	0.28	62.5	34.81
Nitroethane	1.3917	0.02	0.33	46.2	28.29
Nitromethane	1.3817	0.06	0.31	52.6	36.56
o-Dichlorobenzene	1.5515	0.00	0.04	52.7	9.99
o-Nitrotoluene	1.5450	0.00	0.27	59.1	25.67
p-Isopropyltoluene	1.4909	0.00	0.19	38.3	2.23
Perfluorobenzene	1.3777	0.00	0.00	31.7	2.03
Phenyl ether	1.5787	0.00	0.20	38.5	3.73
Pyridine	1.5095	0.00	0.52	52.6	12.98
sec-Butylbenzene	1.4895	0.00	0.16	40.3	2.34
Sulfolane[d]	1.4833	0.00	0.88	77.6	43.96
t-Butylbenzene	1.4927	0.00	0.16	39.8	2.34
Tetrachloroethene	1.5053	0.00	0.00	46.6	2.27
Tetrahydrofuran	1.4050	0.00	0.48	38.0	7.43
Tetralin	1.5413	0.00	0.19	47.7	2.77
Toluene	1.4961	0.00	0.14	40.2	2.38
Tributyl phosphate	1.4224	0.00	1.21	37.4	8.18
Triethylamine	1.4010	0.00	0.79	29.1	2.38
Xylene	1.4995	0.00	0.16	41.4	2.39

[a] See Eq. (17) for descriptor definitions

[b] Chloroform is not listed since chloroform has special parameters, and these descriptors are not used in applying the model in chloroform

[c] Mesitylene

[d] Tetrahydrothiophene-1,1-dioxide

eliminating surface tension coefficients that were small, statistically insignificant, or chemically not warranted, the dependencies and final parameter values shown in Tables 3 and 4 were developed. A blank position in these tables indicates that a particular surface tension coefficient was not included.

The coefficients for the AM1 and PM3 parameterizations are quite similar to each other. This is a result of using the class IV charge model, CM1, to map zero-overlap Mulliken charges to the more physical CM1A and CM1P charges. This mapping results in charges that are more consistent between Hamiltonians than are the original Mulliken charges and leads to both of these models having similar ΔG_{ENP} values and thus similar G_{CDS} values. This is illustrated in Tables 11–25. Of the 41 surface tension coefficients listed in Tables 2 and 3, only one has a different sign in the two models, and both of these values are very near to zero. Roughly a quarter of the surface tensions differ by more than 10 cal mol^{-1} Å^{-2}, but only one differs by more than 12 cal mol^{-1} Å^{-2}. The more positive CS surface tension in the organic SM5.4/AM1 model leads to slightly more negative $\hat{\sigma}_i^{(n)}$ values for all elements except oxygen, but the relative $\hat{\sigma}_i^{(n)}$ values in each model are nearly identical. The biggest difference between models is in the surface tension coefficients for sulfur-related atomic types. However, data in Tables 11, 13, 16, and 22 show that, although the partition between ΔG_{ENP} and G_{CDS} may be different for some sulfur compounds – particularly thiophenol – the overall calculated $\Delta G_{\mathrm{S}}^{\circ}$ are very similar in all cases.

Several of the solvent-dependent surface tension coefficients show encouragingly physical trends.

Table 3. SM5.4/AM1 surface tension coefficients[a] in cal mol^{-1} Å^{-2}

	Z	$\lambda = n$	$\lambda = \alpha$	$\alpha = \beta$
$\hat{\sigma}_Z^{(\lambda)}$	H	−52.03		
	C	−39.79	25.63	10.41
	N	−50.97	−23.10	40.55
	O	−55.63	−25.00	
	F	−42.17		
	S	−79.01	−16.41	17.35
	Cl	−57.49		
	Br	−61.10		
	I	−64.47		
	ZZ'	$\lambda = n$	$\lambda = \alpha$	$\lambda = \beta$
$\hat{\sigma}_{ZZ'}^{(\lambda)}$	H—C	0.76		
	H—N	−6.44		−34.54
	H—O	20.13		−41.58
	H—S	35.85		−0.06
	C—C	−13.89		
	O—C	15.58	42.38	8.28
	O—O	12.90	38.03	−17.07
	N—C	−5.73	−23.01	6.85
	O—N	8.81	77.71	62.15
	S—S	10.68		40.40
$\hat{\sigma}_{ZZ'}^{(2,n)}$	C—C	0.51		
		$\lambda = n$	$\lambda = \gamma$	
$\hat{\sigma}^{CS,\lambda}$		23.56	0.0755	

[a] These are the general organic parameters; all values are changed for chloroform, and $\hat{\sigma}^{CS,\lambda}$ is changed for benzene and toluene

Table 4. SM5.4/PM3 surface tension coefficients[a] in cal mol^{-1} Å^{-2}

	Z	$\lambda = n$	$\lambda = \alpha$	$\lambda = \beta$
$\hat{\sigma}_Z^{(\lambda)}$	H	−47.75		
	C	−37.32	19.75	2.44
	N	−40.56	−11.36	47.75
	O	−59.53	−31.41	
	F	−38.97		
	S	−71.88	−31.08	38.63
	Cl	−54.22		
	Br	−58.88		
	I	−61.89		
	ZZ'	$\lambda = n$	$\lambda = \alpha$	$\lambda = \beta$
$\hat{\sigma}_{ZZ'}^{(\lambda)}$	H—C	−0.80		
	H—N	−6.54		−39.10
	H—O	26.75		−43.73
	H—S	24.10		−40.91
	C—C	−17.81		
	O—C	28.53	56.71	19.23
	O—O	8.77	35.93	−22.79
	N—C	−9.46	−28.36	5.21
	O—N	22.36	64.47	68.76
	S—S	10.71		29.70
$\hat{\sigma}_{ZZ'}^{(2,n)}$	C—C	6.53		
		$\lambda = n$	$\lambda = \gamma$	
$\hat{\sigma}^{CS,\lambda}$		22.07	0.0674	

[a] These are the general organic parameters; all values are changed for chloroform, and $\hat{\sigma}^{CS,\lambda}$ is changed for benzene and toluene

Hydrogens attached to nitrogen and oxygen have negative β surface tension coefficients, meaning that they undergo energetically more favourable interactions with the basic sites on the solvent molecules than are accounted for by electrostatics alone. Similarly, the α components of the surface tensions associated with nitrogen atoms and oxygens in alcoholic solutes are also negative, reflecting first-solvation-shell interactions with hydrogen-bond-donating solvents that are more favourable than the electrostatic component. Interestingly, the $\hat{\sigma}_Z^{(n)}$ results are well correlated with calculated [75] atomic polarizabilities ($R = 0.82$); larger $\hat{\sigma}^{(n)}$ coefficients for more polarizable atoms are expected since components proportional to n were included specifically to capture dispersion-related interactions. For the halogens, which are not affected by atomic types of the form X–Y, the correlation with the calculated atomic polarizabilities is 0.97, and can be fit perfectly $\left(R^2 = 1.000\right)$ with a quadratic fitting scheme. The N—C atomic type was originally added because examination of the experimental data in aqueous solution shows anomalously strong hydrogen bond accepting capabilities for secondary and tertiary amines. In this model the N—C atomic type has a favourable interaction with the hydrogen bond donation ability of the solvent.

The atomic types of the form O–Y all correspond to highly exposed oxygens with large negative charges. While one would expect these atomic types to interact favourably with the solvent, particularly those solvents which can donate hydrogen bonds, the surface tension coefficients for these atomic types are consistently positive. It should be recalled from Eq. (5) that the surface tension coefficients for the atomic types i–j serve to modify the overall surface tension for that atom. Thus, these positive coefficients for highly exposed oxygens indicate that the magnitude of the favourable oxygen-solvent first-solvation-shell effects is reduced. However, the overall surface tension for these exposed oxygens remains negative. While the magnitude of the O—N surface tension coefficients appears to indicate that nitro group oxygens would have an overall positive surface tension, Table 8 shows that solvent descriptor n is 3–5 times greater in magnitude than solvent descriptors α and β. In addition, the maximum values of the distance functions $f_{ZZ'}(r_{kk'})$ and $f_{ZZ'}^{(2)}(r_{kk'})$ in Eq. (11) are approximately 0.7, further reducing the impact of the atomic type X–Y surface tension coefficients. For instance, an oxygen in nitromethane has an overall surface tension of $-32\,\mathrm{cal\,mol}^{-1}$ when dissolved in 1-octanol, while the overall surface tension on the oxygen in methanol dissolved in the same solvent has a surface tension of $-89\,\mathrm{cal\,mol}^{-1}$. Thus the stronger electrostatic interactions created by exposing the negative oxygen to the solvent are partially offset in these models by somewhat weaker, but still favourable, first-solvation-shell effects.

4.2 Special solvents

The parameters in Tables 3 and 4 were obtained by considering all the data in all 90 solvents. These parameters are called the SM5.4-organic parameters. Having obtained these parameters, we examined the quality of the fit solvent-by-solvent. We found only three solvents for which we deemed it worthwhile to further refine the parameters, namely chloroform, benzene, and toluene. The special parameters obtained for these cases are labeled SM5.4-chloroform, SM5.4-benzene, and SM5.4-toluene. Thus the final SM5.4 solvation model involves using the SM5.4-aqueous, SM5.4-chloroform, SM5.4-benzene, and SM5.4-toluene parameters for their respective solvents, and the SM5.4-organic parameters for all other solvents (the 87 other organic solvents used in the SM5.4-non-aqueous solvent training set plus all other organic solvents to which the model can be applied), and these are the parameters we imply when we just say SM5.4 without specifying a parameter set. The parameterization of the SM5.4-chloroform model has been presented in a separate paper [24], and in the rest of this section we discuss the SM5.4-benzene and SM5.4-toluene parameters.

The general SM5.4-organic parameters lead to mean signed errors of 0.5–0.8 kcal/mol for the solvents benzene and toluene. One might consider developing a new set of parameters for those solvents using only data for those solvents. However, such data span only a few solute functional groups such as alcohols, esters, and amines; other functional groups, such as ethers and halogens, are severely underrepresented or non-existent for arene solvents. In contrast, most common functional groups have significant representation in the full non-aqueous training set. Thus, we developed benzene and toluene models based on the organic model similarly to the way we developed a chloroform model [24] based on the organic model. This method develops a new organic model heavily weighted with data from the solvent of interest, but drawing on data from other solutes not represented in the arene under consideration.

For the benzene model, development began with the 1784 solute-solvent data points in the non-aqueous training set, excluding H_2. This set was modified as follows: if a solute was one of the 60 solutes for which the free energy of solvation in benzene was known, the solute-benzene data point was weighted by the number of times that the solute appeared in the non-aqueous training set and all other data points involving that solute were removed. If a solute was not one of the 60 solutes, it was left untouched. This process resulted in a training set with a total weight remaining at 1784. Of that, the 60 benzene data points had a total weight of 1326. Due to the removal of hydrazine from the PM3 set, the total weight in the PM3 model was 1780 and the weight of the benzene points was 1322. The remaining weight involved solutes for which no benzene data were available.

The SM5.4-toluene models were developed in the same way as the benzene models with the exception that toluene free energies of solvation were used instead of benzene free energies. The 45 toluene data points had a weight of 1174 in both the /AM1 and /PM3 models out of total weights of 1784 and 1780, respectively.

Experimentally, for a given solute, ΔG_S° tends to be more negative in benzene than in other arenes and more negative in toluene than in other arenes except benzene. This is typically true even when comparison is made to other arenes that are minimally more substituted, such as xylene and ethylbenzene. The fact that the general organic parameters lead to mean signed errors in both benzene and toluene that are nearly as large as the mean unsigned error in both solvents indicates that the main problem with the general parameters for these two solvents is that they systematically undersolvate nearly all solutes in these solvents. Because of this, it was determined that only a small perturbation of the parameters was needed. Since the CS surface tension applies to all compounds, it was decided simply to modify this term and leave all other surface tension coefficients the same as the organic model. In fact, we found that the systematic error could be removed quite satisfactorily by modifying only the γ-dependent surface tension coefficient of the CS surface tension. Therefore the only modification needed to convert the general organic model to either the benzene or toluene model is

$$\hat{\sigma}^{\mathrm{CS},\lambda}(\text{benzene or toluene}) = \chi\hat{\sigma}^{\mathrm{CS},\lambda}(\text{organic}) \qquad (19)$$

where χ is the appropriate scale factor for the arene model under consideration. (Note that one could alternatively think of this as using effective surface tensions for these two solvents that differ from their true macroscopic surface tensions.) Table 5 gives the optimal scale factors for SM5.4-benzene and SM5.4-toluene. The scale factors were optimized separately for both AM1 and PM3. Because χ is so similar for both Hamiltonians, the values were averaged and rounded, and the scale factor was taken to be the same for both Hamiltonians. Analogous scaling for other arenes did not appear warranted (except possibly for xylene and 1,3,5-trimethylbenzene, but we eventually decided to use scaling only for benzene and toluene). Table 6 shows that the models for benzene and toluene are quite successful.

Table 5. Scale factors in the SM5.4-benzene and -toluene parameterizations

Solvent	Separate values		Final value
	AM1	PM3	
Benzene	0.663	0.652	0.66
Toluene	0.779	0.769	0.77

5 Results and discussion

All solution calculations in this paper fully relaxed both the solute geometry and the electronic wave function in the solvent of interest. All SM5.4 results quoted in this paper for data in the solvents chloroform, benzene, and toluene are calculated with the SM5.4-chloroform, -benzene, and -toluene models, respectively. All energetic quantities are tabulated and discussed on a molar basis, so in the rest of this paper we shorten kcal/mol to kcal.

The range of the solvent properties for each solvent class is given in Table 7. Table 8 gives a summary of the errors in prediction over solvent classes, Table 9 gives an error summary for each organic solvent in the SM5.4-non-aqueous solvent training set, and Table 10 gives an error summary for each solute class. Tables 11–25 give a partial listing of results for selected data used in the model development. Table 26 gives results for some widely measured solutes in a diverse set of solvents.

Table 6. Comparison of errors in free energy of solvation (kcal/mol) predicted by the SM5.4-organic, -benzene, and -toluene parameterizations

Solvent	-organic				-benzene and -toluene			
	AM1		PM3		AM1		PM3	
	MSE[a]	MUS[b]	MSE	MUS	MSE	MUS	MSE	MUS
Benzene	0.8	0.8	0.7	0.7	0.2[c]	0.5[c]	0.2[d]	0.4[d]
Toluene	0.5	0.5	0.5	0.5	0.1[c]	0.3[c]	0.2[d]	0.3[d]

[a] Mean signed error
[b] Mean unsigned error
[c] This is the final value for the SM5.4/AM1 parameterization
[d] This is the final value for the SM5.4/PM3 parameterization

Table 7. Range of solvent properties in each solvent class

Solvent class	n	α	β	γ	ε
Alkanes	1.36–1.48	0.00	0.00	22–44	1.84–2.20
Arenes	1.49–1.54	0.00	0.14–0.19	38–48	2.23–2.77
Alcohols	1.36–1.54	0.33–0.37	0.34–0.56	32–51	7.53–24.9
Ketones	1.38–1.54	0.00	0.48–0.56	33–56	12.9–18.2
Esters	1.37–1.39	0.00	0.45	34–36	4.99–5.99
Ethers	1.35–1.58	0.00	0.20–0.48	24–51	3.05–7.43
Amines	1.40–1.59	0.00–0.26	0.41–0.79	29–61	2.38–6.89
Pyridines	1.50–1.51	0.00	0.52–0.63	45–53	7.17–13.0
Nitriles	1.34–1.53	0.00–0.07	0.32–0.33	41–44	25.6–37.5
Nitrohydrocarbons	1.38–1.56	0.00–0.06	0.27–0.33	46–63	25.7–36.6
Tertiary amides	1.43–1.44	0.00	0.74–0.78	48–51	37.2–37.8
Haloaliphatics	1.39–1.60	0.00–0.15	0.00–0.15	34–65	2.23–10.2
Haloaromatics	1.38–1.62	0.00	0.00–0.12	38–56	2.03–10.0
Misc. acidic solvents	1.37–1.43	0.30–0.61	0.44–0.84	39–56	6.25–181.6
Misc. non-acidic solvents	1.42–1.63	0.00	0.07–1.21	37–78	2.61–46.8

Table 8. Errors in predicted free energies of solvation from the SM5.4 model by solvent functional group class

Solvent class	Number of			SM5.4/AM1		SM5.4/PM3	
	Solute Classes[a]	Solvents[b]	Data[c]	Signed[d]	Unsigned[d]	Signed[d]	Unsigned[d]
Alkanes	28	13	560	−0.1	0.3	−0.1	0.3
Arenes	14	12	253[e]	0.2	0.4	0.2	0.4
Alcohols	26	14	296[e]	−0.1	0.6	−0.1	0.5
Ketones	10	4	35	−0.4	0.5	−0.2	0.4
Esters	8	2	36	0.3	0.4	0.3	0.4
Ethers	17	7	114[e]	0.0	0.5	0.0	0.5
Amines	6	2	12	0.6	0.7	0.5	0.6
Pyridines	5	3	15	0.1	0.3	0.1	0.3
Nitriles	5	2	10	−0.5	0.5	−0.4	0.5
Nitrohydrocarbons	8	4	27	−0.7	0.7	−0.5	0.6
Tertiary amides	5	2	10	−0.1	0.3	−0.1	0.3
Haloaliphatics	24	12	258[e]	0.0	0.6	0.0	0.5
Haloaromatics	11	6	106	−0.3	0.5	−0.2	0.5
Miscellaneous acidic solvents	5	3	15	0.0	0.4	0.0	0.4
Miscellaneous non-acidic solvents	12	4	39	−0.1	0.5	−0.1	0.4

[a] Number of solute classes for which data exist in this solvent class
[b] Number of solvents in this solvent class
[c] Total number of solute/solvent data involving this solvent class
[d] Mean errors in kcal over data in this solvent class
[e] The /PM3 numbers are 1 less due to the removal of hydrazine from the PM3 set

Table 27 gives results for two solvents not included in the SM5.4-non-aqueous solvent training set. Table 28 presents an application of the SM5.4/AM1 model.

5.1 Performance by solvent

Table 8 presents mean errors in free energies of solvation for each solvent class. It shows that the SM5.4 model does well for all kinds of organic solvents. The largest mean unsigned errors are 0.7 kcal in amine and nitro solvents. Very few solvent classes show systematic signed errors; the exceptions are amines, nitriles, and nitro compounds. While the latter three all involve nitrogen-containing solvents, the fact that the signed error is in opposite directions for amines and nitriles indicates that the error is not so much a systematic mistreatment of nitrogen-containing solvents as perhaps a function of the small size of the data sets involved. There is no systematic error for aromatic hydrocarbon solvents as a result of the special attention that we gave to benzene and toluene. Omitting chloroform, benzene, and toluene leaves 1599 data for SM5.4/AM1 and 1597 data for SM5.4/PM3. The mean unsigned error over these data is 0.45 and 0.41 kcal, respectively, which is remarkably successful for a single set of parameters applied to data in 87 solvents.

The performance of the SM5.4 parameterizations by solvent is shown in Table 9. That table shows that, overall, most solvents are treated with about the same accuracy. No solvent with 15 or more measurements has an error above 0.6 kcal mol^{-1}. For organic SM5.4/PM3, only 12 solvents have errors over 0.6 kcal mol^{-1}, and only 2 solvents (both haloalkanes with very limited data sets) have errors greater than 1.0 kcal mol^{-1}. For organic SM5.4/A, there are 16 solvents with errors above 0.6 kcal mol^{-1}, and 3 solvents have errors over 1.0 kcal mol^{-1}.

The dispersion values in Table 9 are measures of the range in magnitude of the experimental data for each solvent. They are calculated as

$$\text{Dispersion} = \left[\overline{\left(\Delta G^\circ_{\text{expt}}\right)^2} - \left(\overline{\Delta G^\circ_{\text{expt}}}\right)^2 \right]^{1/2} \tag{20}$$

One trend that is made clear by Table 9 is the sparsity of data in most solvents. Only a few solvents have enough diverse data to warrant a single-solvent parameterization such as the SM5.4-aqueous parameterizations. More than 90% of the solvents in Table 9 have data sets that are either too sparse or lack enough diversity to create a general single-solvent parameterization. This demonstrates the strength and necessity of the type of parameterizations presented here. We can now predict free energies of solvation in solvents that have very little or no data, such as acetic acid and dimethyl sulfoxide (DMSO).

Table 9. Errors in predicted free energies of solvation from the SM5.4 parameterizations by solvent

	Data[a]	Dispersion[b]	SM5.4/AM1		SM5.4/PM3	
			Signed[c]	Unsigned[c]	Signed[c]	Unsigned[c]
Alkanes						
n-Pentane	26	1.2	−0.3	0.3	−0.3	0.3
n-Hexane	51	1.4	−0.1	0.3	−0.1	0.3
n-Heptane	55	1.5	−0.1	0.3	−0.1	0.3
n-Octane	34	1.2	−0.3	0.3	−0.3	0.3
n-Nonane	25	1.2	−0.2	0.3	−0.2	0.3
n-Decane	36	1.1	−0.2	0.3	−0.2	0.3
n-Undecane	11	0.8	0.0	0.3	0.0	0.2
n-Dodecane	7	1.3	−0.2	0.2	−0.3	0.3
n-Pentadecane	8	0.9	−0.4	0.4	−0.4	0.4
n-Hexadecane	179	2.0	0.1	0.3	0.0	0.3
2,2,4-Trimethylpentane	29	1.1	−0.3	0.4	−0.3	0.4
Cyclohexane	75	1.5	0.0	0.3	0.0	0.3
Decalin	24	1.0	−0.1	0.5	−0.2	0.5
Arenes						
Benzene[d]	60[e]	1.8	0.2	0.5	0.2	0.4
Toluene[d]	45	1.6	0.1	0.3	0.2	0.3
Ethylbenzene	28	1.7	0.2	0.3	0.2	0.3
Isopropylbenzene	17	1.2	0.2	0.4	0.2	0.3
Butylbenzene	9	0.7	0.3	0.4	0.3	0.4
sec-Butylbenzene	5	0.8	−0.1	0.4	−0.1	0.4
t-Butylbenzene	13	0.8	0.1	0.3	0.1	0.3
Xylene	47	1.4	0.4	0.5	0.4	0.5
p-Isopropyltoluene	5	1.1	0.1	0.2	0.1	0.2
1,2,4-Trimethylbenzene	10	0.9	0.3	0.4	0.2	0.3
1,3,5-Trimethylbenzene	6	0.9	0.5	0.5	0.5	0.5
Tetralin	8	1.7	0.8	0.8	0.8	0.8
Alcohols						
Ethanol	5	0.5	−1.0	1.0	−0.9	0.9
1-Propanol	5	0.3	−0.6	0.6	−0.6	0.6
Isopropanol	5	0.2	−0.8	0.8	−0.8	0.8
1-Butanol	18	1.6	0.2	0.5	0.1	0.5
2-Butanol	7	1.6	−0.3	0.6	−0.3	0.5
Isobutanol	16	1.4	−0.1	0.5	−0.1	0.4
1-Pentanol	20	2.2	0.1	0.5	0.2	0.6
1-Hexanol	13	2.4	0.1	0.4	0.2	0.5
1-Heptanol	11	2.3	0.1	0.4	0.2	0.5
1-Octanol	165[e]	2.2	−0.1	0.6	−0.2	0.5
1-Nonanol	9	2.1	0.0	0.4	0.0	0.5
1-Decanol	10	2.2	0.2	0.5	0.2	0.5
Benzyl alcohol	7	1.1	0.4	0.6	0.3	0.6
m-Cresol	5	1.0	0.7	0.8	0.7	0.8

Table 9. Contd.

	Data[a]	Dispersion[b]	SM5.4/AM1		SM5.4/PM3	
			Signed[c]	Unsigned[c]	Signed[c]	Unsigned[c]
Ketones						
Butanone	10	1.8	−0.6	0.6	−0.5	0.5
4-Methyl-2-pentanone	12	2.2	−0.5	0.7	0.0	0.5
Cyclohexanone	7	1.0	0.0	0.3	0.0	0.3
Acetophenone	6	0.7	−0.2	0.2	−0.2	0.2
Esters						
Ethyl acetate	17	1.5	0.2	0.4	0.2	0.4
Butyl acetate	19	2.0	0.3	0.5	0.3	0.5
Ethers						
Ethyl ether	62[e]	2.0	0.0	0.6	0.0	0.5
Tetrahydrofuran	5	0.4	0.2	0.4	0.3	0.4
Isopropyl ether	20	2.3	0.0	0.6	0.1	0.6
Butyl ether	12	0.8	0.3	0.4	0.3	0.4
Anisole	5	0.5	0.3	0.3	0.3	0.3
Ethoxybenzene	5	0.6	0.1	0.1	0.1	0.1
Phenyl ether	5	0.6	−0.5	0.5	−0.5	0.5
Amines						
Triethylamine	5	0.7	0.4	0.4	0.4	0.4
Aniline	7	5.1	0.7	0.8	0.5	0.7
Pyridines						
Pyridine	5	0.3	0.2	0.3	0.2	0.4
2-Methylpyridine	5	0.2	0.1	0.3	0.1	0.3
2,6-Dimethylpyridine	5	0.2	0.1	0.3	0.1	0.3
Nitriles						
Acetonitrile	5	0.6	−0.4	0.4	−0.3	0.4
Benzonitrile	5	0.4	−0.7	0.7	−0.5	0.5
Nitrohydrocarbons						
Nitromethane	5	0.8	−0.2	0.4	−0.1	0.4
Nitroethane	5	0.5	−0.3	0.3	−0.2	0.3
Nitrobenzene	12	1.7	−0.9	0.9	−0.7	0.7
o-Nitrotoluene	5	1.8	−1.0	1.0	−0.9	0.9
Tertiary amides						
N,*N*′-Dimethylacetamide	5	0.5	−0.1	0.3	0.0	0.3
N,*N*′-Dimethylformamide	5	0.5	−0.2	0.4	−0.1	0.3
Haloaliphatics						
1-Fluoroctane	6	0.6	−1.1	1.1	−0.9	0.9
Methylene chloride	9	2.0	−0.6	0.7	−0.2	0.5
Chloroform[f]	82[e]	1.8	0.0	0.4	0.0	0.3
Carbon tetrachloride	67	1.6	0.4	0.5	0.4	0.5
1,2-Dichloroethane	35	1.9	−0.6	0.6	−0.4	0.5
Tetrachloroethene	8	0.9	0.5	0.5	0.4	0.4
1-Chlorohexane	11	0.7	−1.1	1.1	−1.1	1.1
Bromoform	11	1.8	0.7	0.7	0.5	0.5
Bromoethane	5	0.2	−0.2	0.5	0.0	0.5
1,2-Dibromoethane	10	2.1	0.2	0.3	0.1	0.2
1-Bromooctane	5	0.9	−1.6	1.6	−1.5	1.5
1-Iodohexadecane	9	0.6	0.2	0.5	0.2	0.6
Haloaromatics						
Fluorobenzene	5	0.7	−0.1	0.3	0.0	0.3
Chlorobenzene	36	1.6	−0.5	0.5	−0.3	0.4
Bromobenzene	24	1.5	−0.4	0.5	−0.4	0.5
Iodobenzene	19	1.6	−0.3	0.4	−0.3	0.3
o-Dichlorobenzene	10	1.4	−0.7	0.7	−0.7	0.7
Perfluorobenzene	12	0.7	0.7	0.7	0.7	0.7
Micellaneous acidic solvents						
N-Methylformamide	5	0.6	−0.1	0.3	−0.2	0.3
Acetic acid	5	0.6	0.5	0.5	0.4	0.4
2-Methoxyethanol	5	0.4	−0.3	0.4	−0.3	0.3
Misc. non-acidic solvents						
Tributylphosphate	17	1.7	−0.2	0.4	−0.3	0.4
Sulfolane[g]	5	0.8	0.2	0.3	0.1	0.3
Dimethylsulfoxide	5	0.8	0.0	0.3	0.0	0.4
Carbon disulfide	12	1.0	0.0	0.3	0.0	0.3

[a] Number of solutes for which air/*solvent* data were available

[b] Dispersion (kcal) in the experimental data

[c] Mean errors (kcal) over data in this solvent

[d] SM5.4-benzene and SM5.4-toluene results

[e] Number of solutes in the SM5.4/PM3 data set is one less due to the removal of hydrazine

[f] SM5.4-chloroform results

[g] Tetrahydrothiophene-1,1-dioxide

At first inspection, the data in Table 9 for alcohols appear to indicate that there is a systematic error towards overestimating the solvation of molecules in the shorter-chain alcohols. Closer inspection reveals that this is not the case. For instance, in *n*-octanol, the mean signed error for the molecules in the ethanol, propanol, and isopropanol data sets is −0.5 kcal, similar to the errors in propanol and isopropanol. The ethanol data are skewed by what appears to be an incorrect experimental measurement for chlorobenzene. The experimental free energy of solvation of chlorobenzene in ethanol is −3.3 kcal, in *n*-octanol the number is −5.0 kcal, in *n*-decanol it is −4.8 kcal. The average deviation between ethanol and *n*-octanol for the other four compounds in the ethanol data set is 0.2 kcal, and three out of the four have a more negative value in ethanol. If we substitute the experimental chlorobenzene in *n*-octanol value for the experimental chlorobenzene in ethanol value, the error over the ethanol set drops to −0.65 kcal.

A very small systematic error occurs for some of the smaller *n*-alkane solvents. While this error remains within the range of the probable experimental error, which has been estimated by others [16] and independently by ourselves to be about of 0.2 kcal when obtaining data from water/solvent partition coefficients, we do note that we have previously developed a more accurate solvation model which may be used specifically for calculations in alkane solvents [11].

Finally, for several arenes other than benzene and toluene, there is a systematic overestimation of solute free energies of solvation, this being somewhat larger for SM5.4/AM1 than SM5.4/PM3. The magnitude of this effect is typically small, however, and moreover the data in many of these solvents are sparse. Since the error does *not* appear to be significant in xylene (the only arene other than benzene and toluene for which a significant number of measurements is available) we did not consider these solvents to merit further attention.

An important aspect of the SM5 formalism is that first-solvation shell effects are calculated from solvent-accessible surfaces based on well-established van der Waals radii (Bondi's values) without re-optimization and based on a physically reasonable value for the short-range solvent radius (that is, based on considerations of the range of the dispersion force) rather than on numerical re-optimization for each solvent. Although some

Table 10. Errors in predicted free energies of solvation from the SM5.4 models by solute functional group class

Solute class		Number of			SM5.4/AM1		SM5.4/PM3	
		Solutes[a]	Solvent Classes[b]	Data[c]	Signed[d]	Unsigned[d]	Signed[d]	Unsigned[d]
1	Unbranched alkanes	9	15	76	0.2	0.4	0.3	0.4
2	Branched alkanes	5	2	7	0.4	0.4	0.3	0.4
3	Cycloalkanes	4	4	13	1.0	1.0	1.1	1.1
4	Alkenes	8	3	18	0.4	0.5	0.3	0.4
5	Alkynes	5	2	9	−0.1	0.3	−0.1	0.2
6	Arenes	9	15	126	−0.1	0.3	−0.1	0.3
7	Alcohols	17	15	369	0.0	0.4	0.0	0.4
8	Ethers	9	15	71	0.0	0.4	0.0	0.4
9	Aldehydes	7	5	32	0.1	0.5	0.0	0.5
10	Ketones	12	14	191	−0.1	0.4	−0.1	0.4
11	Carboxylic acids	5	11	119	0.3	0.6	0.2	0.6
12	Esters	12	6	227	−0.3	0.5	−0.3	0.5
13	Aliphatic amines	11	8	153	−0.1	0.4	0.0	0.3
14	Aromatic amines	11	10	71	−0.3	0.4	−0.2	0.4
15	Nitriles	4	4	18	−0.1	0.4	0.0	0.4
16	Nitrohydrocarbons	6	6	32	−0.1	0.5	0.0	0.2
17	Ethanamide	1	3	4	−1.1	1.1	−0.8	0.8
18	Thiols	3	3	10	0.1	0.4	0.1	0.3
19	Sulfides	4	3	10	0.0	0.4	−0.1	0.5
20	Disulfides	2	2	3	0.0	0.0	0.0	0.1
21	Non-halo bifunctional solutes	6	6	26	0.3	0.9	0.2	0.9
22	Fluorohydrocarbons	5	3	13	0.4	0.4	0.2	0.3
23	Chlorohydrocarbons	14	4	58	0.0	0.3	0.1	0.3
24	Bromohydrocarbons	10	4	30	−0.1	0.3	−0.2	0.3
25	Iodohydrocarbons	9	4	20	0.0	0.2	0.0	0.2
26	Multifunctional halogenated solutes	13	7	44	0.0	0.6	0.1	0.5
27	Inorganic compounds	5[e]	9	36[e]	−0.7	1.4	−0.6	0.9
	All solutes	206[e]	15	1786[e]	0.0	0.5	−0.1	0.4

[a] Number of solutes in this solute class

[b] Number of solvent classes (as defined in Table 5) for which there are data for this solute class

[c] Total number of solute/solvent data involving solutes in this solute class

[d] Mean errors (kcal) over this solute class data

[e] The SM5.4/PM3 model numbers are 4, 32, 205, and 1782 due to the removal of hydrazine from the PM3 set

improvement in the mean errors could no doubt be obtained by re-optimizing these parameters, the present model retains a more clear physical interpretation than would be applicable if such re-optimization were carried out.

5.2 Performance by solute

Table 10 examines the performance of the organic SM5.4 parameterizations for organic solvents by solute class. As with solvent classes, the overall performance of

Table 11. Selected data and results (kcal) in alkane solvents[a]

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\mathrm{o}}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\mathrm{o}}$	$\Delta G_{\mathrm{S}}^{\mathrm{o}}$
1	*n*-Pentane	Isooctane	−0.1	−2.9	−3.0	0.0	−2.9	−2.9	−3.2
	n-Hexadecane	*n*-Hexadecane	0.1	−10.6	−10.5	0.1	−10.3	−10.2	−10.5
2	2,2-Dimethylpropane	*n*-Hexadecane	−0.1	−1.9	−2.1	−0.1	−2.0	−2.1	−2.5
	2-Methylpentane	*n*-Hexadecane	−0.1	−2.8	−2.8	0.0	−2.8	−2.8	−3.5
3	Cyclopropane	*n*-Hexadecane	−0.4	−0.7	−1.1	−0.2	−1.1	−1.3	−1.8
	Cyclohexane	Cyclohexane	−0.2	−2.5	−2.6	−0.1	−2.5	−2.5	−4.4
4	Propene	Isooctane	−0.4	−1.0	−1.4	−0.2	−1.3	−1.5	−1.6
	s-trans-1,3-Butadiene	*n*-Hexadecane	−0.7	−1.1	−1.8	−0.4	−1.6	−2.0	−2.1
5	Ethyne	*n*-Hexadecane	−1.1	0.9	−0.2	−0.9	0.8	−0.1	−0.2
	1-Pentyne	*n*-Hexadecane	−1.2	−1.8	−3.0	−0.9	−1.9	−2.8	−2.7
6	Benzene	*n*-Decane	−1.4	−2.5	−4.0	−0.9	−3.3	−4.1	−3.8
	Ethylbenzene	*n*-Undecane	−1.3	−3.9	−5.2	−0.8	−4.4	−5.2	−5.4
7	1,2-Ethanediol	*n*-Hexadecane	−3.1	0.1	−3.1	−3.1	−0.2	−3.4	−2.8
	Phenol	Decalin[b]	−3.5	−2.1	−5.6	−2.7	−2.9	−5.6	−5.4
8	1,2-Dimethoxyethane	*n*-Hexadecane	−1.6	−2.3	−3.9	−1.2	−2.6	−3.8	−3.6
	Anisole	*n*-Heptane	−1.9	−3.5	−5.4	−1.3	−4.2	−5.5	−3.9
9	Ethanol	*n*-Hexadecane	−2.1	0.4	−1.7	−2.2	0.4	−1.9	−1.7
	Benzaldehyde	*n*-Hexane	−2.5	−3.1	−5.5	−2.2	−3.6	−5.7	−5.5
10	3,3-Dimethylbutanone	*n*-Heptane	−1.6	−2.6	−4.3	−1.6	−2.6	−4.2	−4.3
	4-Heptanone	*n*-Hexadecane	−1.6	−3.7	−5.3	−1.7	−3.6	−5.3	−5.2
11	Ethanoic acid	Cyclohexane	−3.7	1.4	−2.3	−3.7	1.2	−2.5	−1.7
	Hexanoic acid	*n*-Hexadecane	−3.3	−1.7	−5.0	−3.3	−1.8	−5.1	−5.4
12	Ethyl ethanoate	*n*-Pentane	−2.2	−1.8	−4.0	−2.1	−1.9	−4.0	−3.7
	Methyl pentanoate	*n*-Pentadecane	−2.2	−2.7	−4.9	−2.0	−2.8	−4.9	−4.6
13	Ethylamine	*n*-Hexane	−0.9	−1.9	−2.8	−0.6	−2.0	−2.7	−2.1
	Diethylamine	Cyclohexane	−0.5	−2.9	−3.4	−0.4	−3.3	−3.6	−3.6
14	2-Ethylpyrazine	*n*-Octane	−2.0	−3.9	−5.8	−1.6	−4.4	−6.0	−5.5
	3-Methylpyridine	Cyclohexane	−2.0	−3.1	−5.1	−1.5	−3.8	−5.2	−5.1
15	Ethanonitrile	*n*-Heptane	−2.4	−0.2	−2.6	−2.7	0.2	−2.5	−2.1
	Butanonitrile	*n*-Hexadecane	−2.2	−1.6	−3.8	−2.4	−1.1	−3.5	−3.5
16	Nitrobenzene	*n*-Hexane	−3.4	−2.4	−5.9	−3.1	−3.2	−6.2	−6.1
	2-Methyl-1-nitrobenzene	Cyclohexane	−3.5	−2.8	−6.3	−3.1	−3.5	−6.6	−6.7
17	Ethanamide	*n*-Decane	−3.5	−1.0	−4.5	−3.5	−0.8	−4.3	−2.9
	Ethanamide	*n*-Hexadecane	−3.6	−0.9	−4.5	−3.7	−0.7	−4.3	−3.3
18	1-Propanethiol	Isooctane	−0.5	−3.2	−3.7	−0.4	−3.3	−3.7	−3.8
	Thiophenol	*n*-Hexadecane	−1.6	−4.6	−6.2	−1.0	−5.1	−6.1	−5.6
19	Diethyl sulfide	*n*-Hexadecane	−0.7	−3.6	−4.2	−0.7	−3.6	−4.3	−4.2
	Thioanisole	Decalin[b]	−1.9	−5.1	−7.0	−1.4	−5.7	−7.0	−5.5
20	Dimethyl disulfide	*n*-Hexadecane	−1.1	−3.8	−4.9	−1.3	−3.5	−4.7	−4.8
	Diethyl disulfide	*n*-Hexadecane	−1.0	−4.7	−5.7	−1.3	−4.6	−5.9	−5.7
21	2-Propen-1-ol	*n*-Hexadecane	−2.1	−0.6	−2.7	−1.9	−1.0	−3.0	−2.7
	p-Hydroxybenzaldehyde	*n*-Hexane	−4.2	−2.9	−7.1	−3.8	−3.3	−7.1	−9.2
22	Fluoroethane	*n*-Hexadecane	−0.8	0.5	−0.3	−0.7	0.4	−0.4	−0.8
	Fluorobenzene	*n*-Decane	−1.4	−2.0	−3.4	−1.0	−2.6	−3.6	−3.5
23	Tetrachloroethene	*n*-Undecane	0.0	−4.6	−4.6	0.0	−4.5	−4.5	−4.6
	1,1,1-Trichloroethane	Cyclohexane	−0.6	−3.5	−4.1	−0.5	−3.5	−4.0	−4.1
24	2-Bromopropane	*n*-Hexadecane	−0.7	−2.6	−3.3	−0.7	−2.7	−3.3	−3.3
	p-Dibromobenzene	*n*-Heptane	−0.9	−6.3	−7.2	−0.6	−6.9	−7.5	−7.6
25	Diiodomethane	*n*-Hexadecane	−0.3	−5.0	−5.3	−0.4	−4.8	−5.2	−5.3
	Iodobenzene	*n*-Heptane	−1.1	−5.3	−6.4	−0.7	−5.9	−6.5	−6.3
26	2,2-Dichloro-1,1-difluorethyl methyl ether	*n*-Hexadecane	−1.3	−2.8	−4.1	−1.1	−2.8	−3.8	−3.9
	2,2,2-Trifluorethyl vinyl ether	*n*-Hexadecane	−1.7	−0.4	−2.1	−1.5	−0.6	−2.1	−1.9
27	Ammonia	*n*-Hexadecane	−1.7	0.2	−1.6	−1.0	0.6	−0.4	−0.9
	Hydrogen sulfide	*n*-Hexadecane	−0.5	0.0	−0.5	−0.2	−0.5	−0.7	−0.7

[a] Statistics for the full set of solvents are given in Table 8, and solute class numbers are taken from Table 10

[b] No isomer of decalin specified; solvent properties used in calculations are averaged between *cis* and *trans* isomers

the model is very consistent. For SM5.4/PM3, 23 out of 27 solute classes have errors of 0.6 kcal or less, and only 1 class has an error over the 1.0 kcal limit. The numbers for SM5.4/AM1 are similar, with 23 out of 27 classes having an error of 0.6 kcal or less, and only 2 classes with an error over 1.0 kcal. The statistics for the SM5.4/PM3 parameterization are significantly better than those for the SM5.4/AM1 parameterization for inorganic compounds because of the removal of hydrazine from the PM3 data set. Again, the similar performance of the two parameterizations is demonstrated by the fact that (not including the inorganic compounds which differ in data sets between the two parameterizations) 25 out of 26 mean signed errors and 24 out of 26 mean unsigned errors differ by 0.1 kcal or less between the two parameterizations. No error differs by more than 0.2 kcal between the two parameterizations.

Tables 11–25 present results for one fifth of the training set used to develop the SM5.4 parameterizations for organic solvents. Each individual table involves one solvent class as defined in Table 9. These tables are meant to be as representative as possible. To that end, the data in each table were selected in such a way that each solvent used in developing the model appears at least once and the mean signed and unsigned errors for the molecules in each table is within ±0.1 kcal of the overall results for that solvent class, as given in Table 8. In addition, the solute classes listed in each table (which are defined in Table 10) span the range of solute classes for which data were available in that solvent class. Where possible, two representatives were chosen from each solute class.

5.3 Water versus organic solvents

As mentioned in Sect. 1, water is the solvent most often studied by theoretical methods. It is also probably the most unique solvent. Table 26 shows the uniqueness of water by listing the free energies of solvation of several compounds in a number of organic solvents and also water. The small, polar solvents listed in the table were chosen to provide a variety of solvents with solvent properties close to those of water. The other end of the spectrum is represented by *n*-hexadecane; it is large, non-polar, non-hydrogen bonding, and is the most polarizable *n*-alkane solvent. 1-Octanol is presented as an intermediate solvent. It has hydrogen-bonding ability, a reasonably high dielectric constant, and also has a long aliphatic tail.

For the small polar solute ethanol, electrostatic interactions dominate the solute-solvent interactions and all of the polar solvents have similar solvation energies. The prefactors in Eq. (7) for *n*-hexadecane, acetic acid, and formamide (the weakest, second weakest and

Table 12. Selected data and results (kcal) in aromatic hydrocarbon solvents

Solute class	Solute	Solvent[a]	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\mathrm{o}}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\mathrm{o}}$	$\Delta G_{\mathrm{S}}^{\mathrm{o}}$
1	*n*-Pentane	Benzene	−0.1	−3.0	−3.1	0.0	−3.0	−3.0	−3.0
	n-Hexane	Benzene	−0.1	−3.8	−3.9	0.0	−3.8	−3.8	−3.6
3	Cyclohexane	Benzene	−0.2	−3.0	−3.2	−0.1	−3.0	−3.0	−4.1
6	Benzene	Benzene	−1.6	−2.9	−4.5	−1.0	−3.8	−4.8	−4.6
	Toluene	Xylene	−1.7	−3.0	−4.7	−1.1	−3.8	−4.8	−5.1
7	Ethanol	Tetralin	−2.8	−0.1	−2.7	−2.6	−0.1	−2.7	−1.5
	Phenol	Butylbenzene	−3.7	−2.2	−5.9	−2.9	−3.1	−6.0	−6.8
8	1,4-Dioxane	Toluene	−2.2	−2.1	−4.4	−1.7	−2.8	−4.4	−4.9
	1,4-Dioxane	Xylene	−2.2	−1.7	−4.0	−1.7	−2.4	−4.1	−4.9
10	2-Pentanone	1,3,5-Trimethylbenzene	−2.2	−2.1	−4.2	−2.2	−2.0	−4.2	−4.8
	2-Heptanone	1,2,4-Trimethylbenzene	−2.0	−3.7	−5.8	−2.1	−3.6	−5.7	−6.0
11	Propanoic acid	Isopropylbenzene	−4.0	0.3	−3.6	−3.9	−0.1	−3.7	−4.2
	Pentanoic acid	Xylene	−3.8	−1.2	−5.0	−3.8	−1.3	−5.1	−5.7
12	Propyl ethanoate	*sec*-Butylbenzene	−2.8	−2.1	−4.9	−2.6	−2.3	−4.9	−4.6
	Butyl ethanoate	*t*-Butylbenzene	−2.7	−2.9	−5.7	−2.6	−3.1	−5.6	−5.3
	Pentyl ethanoate	*p*-Isopropyltoluene	−2.6	−3.8	−6.4	−2.4	−3.9	−6.4	−6.0
13	Trimethylamine	Ethylbenzene	−1.2	−1.6	−2.8	−0.4	−2.4	−2.8	−2.6
	Piperidine	Xylene	−0.8	−3.1	−3.9	−0.5	−3.6	−4.1	−5.2
14	Pyridine	Xylene	−2.5	−2.0	−4.6	−1.8	−2.9	−4.7	−5.1
	2-Methylpridine	Benzene	−2.5	−3.4	−5.9	−1.7	−4.1	−5.8	−5.9
16	1-Nitropropane	Toluene	−4.5	−0.5	−5.0	−3.8	−0.8	−4.6	−5.3
	Nitrobenzene	Benzene	−4.2	−2.2	−6.4	−3.7	−3.1	−6.8	−7.6
21	*m*-Hydroxybenzaldehyde	Benzene	−4.9	−3.2	−8.0	−4.3	−3.7	−8.1	−9.3
26	*p*-Bromophenol	Ethylbenzene	−3.6	−4.2	−7.8	−2.9	−4.9	−7.9	−8.5
	P-Bromophenol	Xylene	−3.6	−4.2	−7.8	−2.9	−5.0	−7.8	−8.7
27	Water	Benzene	−4.4	2.0	−2.4	−4.1	2.0	−2.1	−1.7
	Water	Toluene	−4.5	2.1	−2.4	−4.3	2.1	−2.2	−1.7

[a] Benzene results are calculated with SM5.4-benzene parameters, and toluene results are calculated with SM5.4-toluene parameters. No isomer is specified for xylene; solvent properties used in calculations are averaged between *o*-, *m*-, and *p*-isomers

strongest dielectric solvents) are 0.26, 0.42, and 0.50 respectively, showing that all solvents but *n*-hexadecane will have reasonably similar electrostatic interactions.

For the somewhat larger solutes, butanone and 1,4-dioxane, the electrostatic interactions with the solvent have decreased, and the magnitudes of the first-solvation-shell effects such as dispersion are beginning to increase. Thus, the solvation energies for these two solutes are more similar in *n*-hexadecane and in water than they are for ethanol. For both of these solutes, the small polar organic solvents behave at least as much like 1-octanol as they do like water.

The non-hydrogen-bonding solutes, *n*-octane and toluene, have dramatically different solvation energies in water than in the organic solvents. For these solutes, several of the small, polar, organic solvents exhibit behavior much more similar to *n*-hexadecane than to water. It is likely that much of this arises from the cost of reordering the solvent to place the non-hydrogen bonding solute into the solvent structure. In water, the

Table 13. Selected data and results (kcal) in alcohol solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	$\Delta G_{\mathrm{S}}^{\circ}$
1	Ethane	1-Octanol	−0.2	0.2	0.0	−0.1	0.0	−0.1	−0.6
	n-Octane	Ethanol	−0.1	−4.4	−4.5	0.0	−4.4	−4.4	−4.2
2	2-Methylpropane	1-Octanol	−0.2	−1.1	−1.4	−0.1	−1.3	−1.4	−1.5
	2,2-Dimethylpropane	1-Octanol	−0.2	−1.6	−1.8	−0.1	−1.8	−1.9	−1.7
3	Cyclopentane	1-Octanol	−0.5	−1.6	−2.1	−0.2	−1.8	−2.0	−2.7
	Methylcyclohexane	1-Octanol	−0.2	−2.8	−3.0	−0.1	−2.8	−2.9	−3.2
4	Propene	1-Octanol	−0.8	0.2	−0.6	−0.4	−0.4	−0.8	−1.1
	2-methylpropene	1-Octanol	−0.9	−0.5	−1.4	−0.5	−1.0	−1.6	−2.0
5	Propyne	1-Octanol	−2.6	1.0	−1.6	−2.1	0.3	−1.8	−1.6
	1-Hexyne	1-Octanol	−2.2	−1.5	−3.7	−1.8	−2.0	−3.7	−3.4
6	Toluene	1-Propanol	−3.0	−1.9	−4.9	−1.9	−3.0	−4.9	−4.5
	Toluene	Isopropanol	−3.0	−2.1	−5.1	−1.9	−3.2	−5.0	−4.4
7	Ethanol	Benzyl alcohol	−4.2	−0.4	−4.6	−3.9	−0.8	−4.7	−4.8
	1,2-Ethanediol	1-Butanol	−6.2	−1.5	−7.6	−6.2	−2.0	−8.2	−8.7
8	1,4-Dioxane	*m*-Cresol	−3.7	−1.9	−5.5	−2.7	−2.9	−5.6	−6.8
	1,2-Dimethoxyethane	1-Octanol	−3.0	−2.3	−5.3	−2.2	−2.8	−5.0	−4.6
9	Methanal	2-Butanol	−4.7	2.3	−2.4	−4.9	2.5	−2.4	−2.9
	Butanal	Isobutanol	−4.0	−1.0	−5.0	−4.2	−0.8	−5.1	−4.8
10	Butanone	Benzyl alcohol	−4.2	−0.7	−4.8	−4.2	−0.7	−4.9	−4.6
	Cyclopentanone	1-Octanol	−3.8	−1.2	−5.0	−4.0	−1.0	−5.0	−5.0
11	Ethanoic acid	1-Heptanol	−7.5	1.6	−5.9	−7.3	−1.3	−6.0	−6.7
	Butanoic acid	1-Pentanol	−6.8	−0.3	−7.1	−6.6	−0.5	−7.1	−7.7
12	Ethyl ethanoate	Isobutanol	−5.3	−0.9	−6.2	−4.9	−1.1	−6.0	−4.3
	Propyl ethanoate	1-Octanol	−4.8	−1.5	−6.3	−4.5	−1.7	−6.1	−4.6
13	Ethylamine	1-Hexanol	−1.9	−2.3	−4.2	−1.3	−2.6	−4.0	−4.2
	Butylamine	1-Decanol	−1.6	−3.7	−5.3	−1.2	−4.0	−5.1	−5.2
14	2-Methylpyrazine	1-Octanol	−5.0	−2.3	−7.2	−3.5	−3.3	−6.8	−5.9
	Aniline	1-Pentanol	−4.5	−2.9	−7.4	−4.0	−3.8	−7.8	−6.4
15	Propanonitrile	1-Octanol	−4.7	0.3	−4.4	−5.3	1.1	−4.2	−3.7
	Benzonitrile	1-Octanol	−4.5	−1.1	−5.6	−4.3	−1.2	−5.5	−6.1
16	1-Nitrobutane	1-Octanol	−7.6	2.0	−5.6	−6.3	1.1	−5.2	−5.1
	2-Methyl-1-nitrobenzene	1-Octanol	−7.1	1.0	−6.1	−6.2	−0.7	−6.8	−6.8
18	1-Propanethiol	1-Octanol	−1.0	−2.4	−3.4	−0.8	−2.8	−3.5	−3.5
	Thiophenol	1-Octanol	−3.1	−3.3	−6.4	−1.9	−4.4	−6.3	−6.0
19	Diethyl sulfide	1-Octanol	−1.3	−3.3	−4.6	−1.4	−3.2	−4.6	−4.1
	Thioanisole	1-Octanol	−3.5	−3.9	−7.4	−2.7	−4.6	−7.3	−6.5
20	Dimethyl disulfide	1-Octanol	−2.1	−2.1	−4.2	−2.5	−1.8	−4.2	−4.2
	2-Methoxyethanol	1-Nonanol	−4.8	−1.9	−6.7	−4.4	−2.5	−6.9	−5.6
21	Morpholine	1-Octanol	−3.3	−3.5	−6.9	−2.6	−4.5	−7.0	−6.0
22	1,1-Difluoroethane	1-Octanol	−2.5	1.0	−1.5	−2.3	0.9	−1.4	−1.1
	Fluorobenzene	1-Octanol	−2.7	−0.7	−3.4	−2.0	−1.8	−3.8	−3.9
23	Dichloromethane	1-Octanol	−1.6	−1.7	−3.3	−1.6	−1.8	−3.4	−3.1
	Chlorobenzene	1-Decanol	−2.4	−2.3	−4.7	−1.5	−3.4	−4.9	−4.8
24	Dibromomethane	1-Octanol	−1.0	−3.2	−4.3	−1.0	−3.3	−4.3	−4.2
	Bromoethane	1-Octanol	−1.3	−1.9	−3.2	−1.3	−2.1	−3.4	−2.9
25	Iodoethane	1-Octanol	−1.0	−2.8	−3.8	−0.8	−2.9	−3.6	−3.5
	2-Iodopropane	1-Octanol	−1.0	−3.2	−4.2	−0.9	−3.3	−4.1	−4.4
26	*p*−Bromophenol	1-Pentanol	−6.3	−3.7	−10.0	−5.0	−4.8	−9.8	−10.6
	1,1,1,3,3,3-Hexafluoropropan-2-ol	1-Octanol	−5.4	0.6	−4.9	−5.2	0.5	−4.7	−5.8
27	Ammonia	1-Pentanol	−3.6	−1.0	−4.6	−2.0	−0.4	−2.4	−3.1
	Water	2-Butanol	−7.8	0.4	−7.4	−7.3	0.2	−7.2	−5.7

Table 14. Selected data and results (kcal) in ketone solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\text{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\text{S}}^{\circ}$	$\Delta G_{\text{S}}^{\circ}$
1	*n*-Octane	Cyclohexanone	−0.1	−4.0	−4.1	0.0	−4.0	−4.1	−4.6
	n-Octane	Acetophenone	−0.1	−4.1	−4.1	0.0	−4.1	−4.1	−4.2
6	Toluene	Butanone	−3.0	−2.6	−5.6	−1.9	−3.5	−5.4	−5.1
	Naphthalene	4-Methyl-2-pentanone	−4.6	−4.0	−8.6	−2.9	−5.5	−8.4	−7.5
7	Ethanol	Cyclohexanone	−4.3	−0.1	−4.4	−4.0	−0.4	−4.4	−4.4
	m-Cresol	4-Methyl-2-pentanone	−6.4	−3.0	−9.4	−5.0	−3.8	−8.9	−8.8
8	1,4-Dioxane	Cyclohexanone	−3.8	−1.4	−5.1	−2.8	−2.1	−4.9	−5.0
	1,4-Dioxane	Acetophenone	−3.8	−1.3	−5.1	−2.8	−2.1	−4.9	−5.0
9	Methanal	Butanone	−4.7	2.0	−2.7	−5.0	2.2	−2.8	−1.8
10	Butanone	Cyclohexanone	−4.3	−0.7	−5.0	−4.3	−0.7	−5.0	−4.4
	Butanone	Acetophenone	−4.3	−0.7	−5.0	−4.4	−0.7	−5.0	−4.4
11	Ethanoic acid	Butanone	−7.8	0.7	−7.1	−7.6	0.5	−7.1	−6.9
	Butanoic acid	4-Methyl-2-pentanone	−6.7	−1.0	−7.7	−6.5	−1.2	−7.7	−7.4
13	Methylamine	4-Methyl-2-pentanone	−2.6	−1.0	−3.6	−1.6	−1.3	−2.9	−4.1
	Diethylamine	4-Methyl-2-pentanone	−1.1	−2.5	−3.6	−0.7	−3.0	−3.7	−3.6
14	Pyridine	4-Methyl-2-pentanone	−5.1	−1.2	−6.3	−3.1	−2.2	−5.3	−5.3
	Aniline	4-Methyl-2-pentanone	−4.4	−3.4	−7.8	−3.9	−4.1	−8.1	−7.5
27	Ammonia	4-Methyl-2-pentanone	−3.6	−0.8	−4.3	−2.0	−0.4	−2.3	−2.5

Table 15. Selected data and results (kcal) in ester solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\text{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\text{S}}^{\circ}$	$\Delta G_{\text{S}}^{\circ}$
1	*n*-Octane	Ethyl acetate	−0.1	−4.5	−4.6	0.0	−4.5	−4.5	−4.7
6	Toluene	Ethyl acetate	−2.6	−2.7	−5.2	−1.6	−3.6	−5.2	−5.1
	Naphthalene	Butyl acetate	−3.9	−4.0	−7.8	−2.4	−5.4	−7.8	−7.6
7	1-Butanol	Ethyl acetate	−3.1	−2.0	−5.1	−3.0	−2.2	−5.2	−5.8
	1-Octanol	Ethyl acetate	−2.9	−5.0	−7.9	−3.0	−5.0	−8.0	−8.4
11	Ethanoic acid	Butyl acetate	−6.4	0.8	−5.5	−6.3	0.6	−5.6	−6.1
	Propanoic acid	Ethyl acetate	−6.1	−0.2	−6.3	−6.0	−0.3	−6.3	−7.0
14	Pyridine	Butyl acetate	−4.0	−1.2	−5.2	−2.6	−2.2	−4.8	−5.3
	Aniline	Butyl acetate	−3.7	−3.3	−7.0	−3.3	−4.0	−7.3	−7.3
26	*p*-Bromophenol	Butyl acetate	−5.2	−4.0	−9.3	−4.2	−4.9	−9.1	−10.6
27	Water	Ethyl acetate	−6.8	1.2	−5.6	−6.4	1.1	−5.3	−4.3
	Water	Butyl acetate	−6.5	1.3	−5.2	−6.1	1.2	−4.9	−4.1

hydrogen bonding network is extensive, and insertion of a non-polar solute either disrupts this network or causes significant reorganization among the solvent molecules to continue the network. Of course, for *n*-hexadecane there is no such issue. For solvents such as the alcohols and acetic acid, it is likely that the hydrogen bond network is much more localized between clusters of molecules with significant portions of the bulk solvent occupied by non-hydrogen bonding aliphatic groups [81–83]. The reorganization cost for placing a solute in these aliphatic regions is probably quite small and similar to that of *n*-hexadecane. It can be argued that the hydrogen-bonding network in 1,2-ethanediol and formamide is probably more extensive than in the alcohols and acid. In the diol, both 'ends' of the molecule can participate in hydrogen bonds, allowing the entire molecule to be part of the network. This is also true to some extent in formamide, where the aliphatic hydrogen has a CM1A charge of about 0.2 in solution, polar enough to participate in weak hydrogen bonding. These two solvents exhibit solvation free energies for toluene that are approximately halfway between *n*-hexadecane and water. Interestingly, the macroscopic surface tension of these liquids is significantly higher than that of *n*-hexadecane and the similar organic solvents, but still significantly lower than that of water. Table 27 shows that the organic SM5.4 parameterizations are able to reproduce this trend for formamide and 1,2-ethanediol quite well *despite the fact that they were not part of the training set for these parameterizations.*

5.4 Case study: solvation of ethanol

The present model specifically allows exploration of physical effects associated with the solvent dependence of solvation energies. As an example, Table 28 lists the solvation free energies of ethanol in a number of organic solvents. Also listed in Table 28 are the solvent properties used in Eqs. (13)–(16), the value of ε, the function of ε that occurs in the electrostatic term, and the calculated free energy of solvation. The last row shows the correlation $\left(R^2\right)$ of these seven columns with the experimental free energy of solvation. Examination

Table 16. Selected data and results (kcal) in ether solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	$\Delta G_{\mathrm{S}}^{\circ}$
1	*n*-Octane	Butyl ether	−0.1	−4.7	−4.7	0.0	−4.7	−4.7	−5.2
	n-Octane	Anisole	−0.1	−4.3	−4.4	0.0	−4.3	−4.3	−4.6
6	Toluene	Tetrahydrofuran	−2.7	−2.6	−5.3	−1.7	−3.5	−5.2	−5.5
	m-Xylene	Ethyl ether	−2.4	−3.7	−6.1	−1.5	−4.4	−5.9	−5.6
7	Ethanol	Phenyl ether	−3.2	−0.3	−3.5	−3.0	−0.5	−3.5	−3.2
	Cyclopentanol	Ethyl ether	−2.4	−2.5	−4.9	−2.3	−2.7	−5.0	−6.5
8	1,4-Dioxane	Isopropyl ether	−2.9	−2.1	−5.0	−2.1	−2.7	−4.9	−4.4
	1,4-Dioxane	Tetrahydrofuran	−3.5	−1.7	−5.2	−2.6	−2.4	−4.9	−5.2
9	Methanal	Isopropyl ether	−3.4	1.7	−1.7	−3.6	1.8	−1.8	−1.0
	Propanal	Ethyl ether	−3.2	−0.7	−3.9	−3.4	−0.6	−3.9	−3.9
10	Butanone	Ethoxybenzene	−3.2	−1.0	−4.2	−3.3	−1.0	−4.3	−4.3
	Methyl phenyl ketone	Ethyl ether	−4.4	−3.4	−7.8	−3.8	−4.1	−7.9	−6.8
11	Ethanoic acid	Butyl ether	−5.2	0.7	−4.5	−5.1	0.5	−4.6	−5.2
	Hexanoic acid	Isopropyl ether	−5.0	−2.8	−7.8	−5.0	−2.8	−7.7	−8.2
13	Trimethylamine	Isopropyl ether	−1.5	−1.6	−3.2	−0.5	−2.3	−2.8	−2.7
	Diethylamine	Butyl ether	−0.7	−2.6	−3.3	−0.5	−3.0	−3.6	−3.8
14	2-Methylpyridine	Butyl ether	−3.2	−2.2	−5.4	−2.1	−3.1	−5.2	−5.2
	2-Methylpyrazine	Butyl ether	−3.3	−1.4	−4.8	−2.5	−2.2	−4.7	−5.1
15	Ethanonitrile	Ethyl ether	−4.3	1.0	−3.3	−4.8	1.5	−3.3	−3.6
	Benzonitrile	Ethyl ether	−3.7	−1.9	−5.6	−3.5	−2.1	−5.5	−6.4
16	Nitrobenzene	Ethyl ether	−6.1	−0.9	−7.1	−5.4	−1.9	−7.2	−6.9
	2-Methyl-1-nitrobenzene	Ethyl ether	−5.7	−1.8	−7.5	−5.0	−2.6	−7.6	−7.2
21	2-Propen-1-ol	Ethyl ether	−3.1	−1.2	−4.3	−2.8	−1.6	−4.4	−4.9
	2-Methoxyethanol	Ethyl ether	−4.1	−1.9	−6.0	−3.8	−2.2	−6.0	−5.1
23	Chlorobenzene	Ethyl ether	−2.0	−3.5	−5.6	−1.3	−4.3	−5.6	−5.4
	p-Dichlorobenzene	Ethyl ether	−1.5	−4.7	−6.2	−1.0	−5.3	−6.2	−6.2
24	Bromobenzene	Ethyl ether	−2.1	−4.2	−6.3	−1.3	−5.1	−6.4	−6.0
25	Iodomethane	Ethyl ether	−0.8	−2.4	−3.2	−0.6	−2.5	−3.1	−3.5
	Iodobenzene	Ethyl ether	−1.9	−5.0	−6.9	−1.2	−5.8	−6.9	−6.6
27	Water	Ethyl ether	−6.2	1.1	−5.1	−5.8	1.0	−4.8	−3.9
	Hydrogen sulfide	Ethyl ether	−0.8	0.3	−0.5	−0.2	−0.2	−0.4	−0.6

Table 17. Complete data and results (kcal) in amine solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	$\Delta G_{\mathrm{S}}^{\circ}$
1	*n*-Octane	Triethylamine	0.0	−4.8	−4.9	0.0	−4.8	−4.8	−5.6
	n-Octane	Aniline	−0.1	−4.0	−4.0	0.0	−4.1	−4.1	−3.5
6	Toluene	Triethylamine	−1.7	−2.7	−4.4	−1.1	−3.8	−4.8	−5.0
	Toluene	Aniline	−2.7	−1.7	−4.4	−1.7	−2.9	−4.6	−4.6
7	Ethanol	Triethylamine	−2.5	−1.0	−3.5	−2.3	−1.2	−3.5	−3.9
	Ethanol	Aniline	−3.9	0.1	−3.7	−3.6	−0.3	−3.8	−4.5
8	1,4-Dioxane	Triethylamine	−2.2	−2.1	−4.3	−1.6	−2.7	−4.3	−4.4
	1,4-Dioxane	Aniline	−3.4	−1.3	−4.7	−2.6	−2.2	−4.8	−5.7
10	Butanone	Triethylamine	−2.3	−1.3	−3.6	−2.3	−1.2	−3.5	−3.9
	Butanone	Aniline	−3.8	−0.5	−4.2	−3.8	−0.5	−4.3	−4.9
11	Ethanoic acid	Aniline	−6.9	2.2	−4.6	−6.7	1.9	−4.8	−6.3
	Propanoic acid	Aniline	−6.3	1.2	−5.2	−6.1	0.8	−5.3	−6.2

of the trends in the table illustrates the especially critical roles of solvent basicity and electrostatic effects in the organic solvation of ethanol.

5.5 Comparison to other models

The only other continuum solvation models parameterized to predict free energies of solvation in organic solvents appear to be the chloroform [13] and carbon tetrachloride [14] models of Luque et al. These models employ the Miertus-Scrocco-Tomasi (MST) approach [84] for calculating ΔG_{ENP} and an approach similar to our own [parameterized atomic surface tension (ST)] for evaluating G_{CDS}. For chloroform, we have previously made [24] a thorough comparison of SM5.4 to the MST/ST model of Luque et al. Table 29 provides a comparison of SM5.4 to the MST/ST model of Luque

Table 18. Selected data and results (kcal) in pyridine solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	$\Delta G_{\mathrm{S}}^{\circ}$
1	*n*-Octane	Pyridine	−0.1	−4.1	−4.2	0.0	−4.2	−4.2	−4.5
	n-Octane	2-Methylpyridine	−0.1	−4.3	−4.4	0.0	−4.4	−4.4	−4.7
6	Toluene	2,6-Dimethylpyridine	−2.7	−2.5	−5.1	−1.7	−3.6	−5.2	−5.0
	Toluene	Pyridine	−2.9	−2.3	−5.2	−1.8	−3.3	−5.1	−5.1
7	Ethanol	2-Methylpyridine	−4.1	−0.2	−4.3	−3.8	−0.5	−4.3	−5.0
	Ethanol	2,6-Dimethylpyridine	−3.9	−0.4	−4.3	−3.6	−0.7	−4.3	−4.9
8	1,4-Dioxane	Pyridine	−3.7	−1.4	−5.1	−2.7	−2.1	−4.9	−5.1
	1,4-Dioxane	2-Methylpyridine	−3.6	−1.5	−5.1	−2.7	−2.3	−5.0	−5.0
10	Butanone	2,6-Dimethylpyridine	−3.8	−1.0	−4.8	−3.8	−0.9	−4.8	−4.3
	Butanone	Pyridine	−4.2	−0.7	−4.9	−4.2	−0.7	−5.0	−4.6

Table 19. Completed data and results (kcal) in nitrile solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	$\Delta G_{\mathrm{S}}^{\circ}$
1	*n*-Octane	Acetonitrile	−0.1	−4.0	−4.0	0.0	−4.0	−4.0	−3.6
	n-Octane	Benzonitrile	−0.1	−4.7	−4.8	0.0	−4.7	−4.7	−4.3
6	Toluene	Acetonitrile	−3.1	−2.2	−5.3	−2.0	−3.1	−5.1	−4.7
	Toluene	Benzonitrile	−3.1	−2.8	−5.9	−1.9	−3.7	−5.7	−5.0
7	Ethanol	Acetonitrile	−4.5	−0.1	−4.6	−4.2	−0.4	−4.6	−4.4
	Ethanol	Benzonitrile	−4.4	−0.1	−4.6	−4.1	−0.4	−4.5	−4.1
8	1,4-Dioxane	Acetonitrile	−4.0	−1.4	−5.4	−3.0	−2.2	−5.1	−5.3
	1,4-Dioxane	Benzonitrile	−3.9	−1.7	−5.6	−2.9	−2.5	−5.3	−5.1
10	Butanone	Acetonitrile	−4.5	−0.9	−5.4	−4.6	−0.9	−5.5	−4.7
	Butanone	Benzonitrile	−4.5	−1.2	−5.6	−4.5	−1.2	−5.7	−4.6

Table 20. Selected data and results (kcal) in solvents containing nitro groups

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	$\Delta G_{\mathrm{S}}^{\circ}$
1	*n*-Octane	Nitromethane	−0.1	−3.5	−3.6	0.0	−3.6	−3.6	−3.2
	n-Octane	Nitroethane	−0.1	−3.9	−4.0	0.0	−4.0	−4.0	−3.9
6	Toluene	Nitromethane	−3.1	−1.9	−5.0	−1.9	−2.8	−4.7	−4.5
	Toluene	Nitroethane	−3.1	−2.3	−5.3	−1.9	−3.1	−5.1	−4.9
7	Phenol	*o*-Nitrotoluene	−6.8	−1.7	−8.4	−5.3	−2.7	−8.0	−7.8
	p-Cresol	Nitrobenzene	−6.9	−2.2	−9.0	−5.3	−3.1	−8.5	−8.1
8	1,4-Dioxane	Nitromethane	−4.0	−1.1	−5.0	−2.9	−1.8	−4.8	−5.5
	1,4-Dioxane	Nitroethane	−3.9	−1.4	−5.3	−2.9	−2.1	−4.9	−5.3
10	Butanone	Nitromethane	−4.5	−0.5	−5.1	−4.6	−0.6	−5.1	−4.7
	Butanone	Nitroethane	−4.5	−0.8	−5.3	−4.5	−0.8	−5.3	−4.7
11	Ethanoic acid	Nitrobenzene	−8.1	2.0	−6.1	−7.9	1.7	−6.3	−4.8
	Butanoic acid	*o*-Nitrotoluene	−7.0	0.2	−6.9	−6.9	−0.1	−6.9	−5.8
14	Aniline	Nitrobenzene	−4.7	−2.8	−7.5	−4.2	−3.6	−7.8	−7.2
26	*p*-Bromophenol	Nitrobenzene	−6.7	−3.5	−10.1	−5.3	−4.4	−9.8	−9.8
	p-Bromophenol	*o*-Nitrotoluene	−6.6	−3.6	−10.2	−5.2	−4.5	−9.8	−9.6

et al. [14] for carbon tetrachloride. For the 27 molecules that Luque et al. used to parameterize their model (which are all of the molecules for which they report free energies of solvation in carbon tetrachloride [14]), the SM5.4/AM1 and SM5.4/PM3 models give mean unsigned errors of 0.5 and 0.4 kcal respectively. Two of these molecules, *o*- and *p*-nitrophenol, were not part of the SM5.4-non-aqueous training set. The performance of the SM5.4 model for the 27 solutes in Table 29 is equivalent to the performance of the model for the full set of 67 CCl_4 data points that are found in the non-aqueous training set (see Tale 9). The AM1- and PM3-based models of Luque et al. give mean unsigned errors of 0.3 kcal/mol for their training set, as does a HF/6-31G* implementation of their model.

A significant difference between the two models, however, is that for the data in Table 29, Luque et al. report ΔG_{ENP} values in the range of −0.2 to −1.8 kcal, whereas the SM5.4 models predict ΔG_{ENP} values in the range of −0.4 to −6.4 kcal, i.e., the SM5.4 models attrib-

Table 21. Selected data and results (kcal) in tertiary amide solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	$\Delta G_{\mathrm{S}}^{\circ}$
1	*n*-Octane	*N*,*N*′-Dimethylformamide	−0.1	−3.8	−3.9	0.0	−3.9	−3.9	−3.8
	n-Octane	*N*,*N*′-Dimethylacetamide	−0.1	−4.0	−4.1	0.0	−4.1	−4.1	−3.9
6	Toluene	*N*,*N*′-Dimethylformamide	−3.1	−1.9	−5.0	−1.9	−3.1	−5.0	−4.9
	Toluene	*N*,*N*′-Dimethylacetamide	−3.1	−2.0	−5.1	−1.9	−3.2	−5.1	−4.9
7	Ethanol	*N*,*N*′-Dimethylformamide	−4.5	−0.2	−4.7	−4.2	−0.5	−4.7	−5.2
	Ethanol	*N*,*N*′-Dimethylacetamide	−4.5	−0.4	−4.9	−4.2	−0.7	−4.8	−5.4
8	1,4-Dioxane	*N*,*N*′-Dimethylformamide	−4.0	−1.3	−5.2	−3.0	−2.0	−5.0	−5.0
	1,4-Dioxane	*N*,*N*′-Dimethylacetamide	−4.0	−1.4	−5.4	−2.9	−2.1	−5.1	−5.0
10	Butanone	*N*,*N*′-Dimethylformamide	−4.5	−0.6	−5.1	−4.6	−0.5	−5.1	−4.6
	Butanone	*N*,*N*′-Dimethylacetamide	−4.5	−0.7	−5.2	−4.6	−0.6	−5.2	−4.5

Table 22. Selected data and results (kcal) in haloaliphatic solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	$\Delta G_{\mathrm{S}}^{\circ}$
1	*n*-Hexane	1-Iodohexadecane	−0.1	−2.8	−3.0	0.0	−2.9	−2.9	−3.3
	n-Octane	Methylene chloride	−0.1	−4.5	−4.5	0.0	−4.4	−4.4	−5.2
3	Cyclohexane	Chloroform	−0.3	−3.1	−3.4	−0.1	−3.1	−3.2	−4.5
	Methylcyclohexane	1-Iodohexadecane	−0.2	−2.7	−2.9	−0.1	−2.8	−2.8	−4.1
4	2-Methylpropene	Carbon tetrachloride	−0.6	−1.3	−1.8	−0.3	−1.5	−1.8	−2.6
	E-2-Pentene	Carbon tetrachloride	−0.5	−2.4	−2.8	−0.3	−2.5	−2.8	−3.5
6	Toluene	Bromoethane	−2.8	−3.1	−5.9	−1.7	−3.8	−5.5	−5.6
	o-Xylene	Chloroform	−2.5	−3.5	−6.0	−1.6	−4.4	−5.9	−6.2
7	Methanol	1,2-Dibromoethane	−3.6	1.4	−2.2	−3.4	1.0	−2.3	−2.4
	Phenol	Tetrachloroethene	−3.6	−2.1	−5.7	−2.8	−2.9	−5.7	−6.1
8	1,4-Dioxane	Bromoethane	−3.6	−1.9	−5.4	−2.7	−2.5	−5.1	−5.4
	Diethyl ether	Chloroform	−1.7	−2.7	−4.4	−1.2	−2.9	−4.2	−4.3
9	Ethanal	Chloroform	−3.6	0.3	−3.3	−3.8	0.4	−3.4	−3.7
	Benzaldehyde	1,2-Dichloroethane	−5.4	−2.0	−7.5	−4.7	−2.8	−7.5	−7.2
10	Propanone	Tetrachloroethene	−2.4	−0.3	−2.7	−2.4	−0.4	−2.8	−3.1
	2-Pentanone	1-Chlorohexane	−3.4	−2.1	−5.5	−3.5	−2.1	−5.6	−4.8
11	Ethanoic acid	Bromoform	−6.0	2.6	−3.4	−5.9	2.3	−3.6	−4.5
	Butanoic acid	Chloroform	−5.4	−0.3	−5.7	−5.3	−0.5	−5.8	−6.0
12	Methyl ethanoate	Tetrachloroethene	−2.9	0.1	−2.8	−2.6	−0.3	−2.9	−3.6
	Propl ethanoate	Bromooctane	−4.2	−2.0	−6.1	−3.9	−2.2	−6.0	−4.5
13	Ethylamine	1,2-Dichloroethane	−1.8	−1.5	−3.3	−1.3	−1.8	−3.1	−3.2
	Dimethylamine	Chloroform	−1.6	−1.8	−3.4	−1.0	−2.6	−3.5	−3.7
14	3-Methylpyridine	Chloroform	−3.8	−3.4	−7.1	−2.5	−4.3	−6.8	−7.4
	Aniline	Carbon tetrachloride	−2.5	−3.5	−6.0	−2.2	−4.0	−6.2	−6.1
15	Ethanonitrile	Chloroform	−4.5	−0.2	−4.6	−5.0	0.3	−4.7	−4.4
	Benzonitrile	Carbon tetrachloride	−2.5	−3.3	−5.8	−2.3	−3.4	−5.7	−6.3
16	1-Nitropropane	Carbon tetrachloride	−4.2	−0.6	−4.8	−3.6	−0.9	−4.5	−4.5
	Nitrobenzene	Chloroform	−6.4	−1.3	−7.7	−5.6	−2.2	−7.7	−7.8
17	Ethanamide	Choloroform	−6.1	−1.3	−7.4	−6.2	−0.9	−7.1	−7.1
18	Thiophenol	Methylene chloride	−3.1	−4.3	−7.4	−1.9	−5.0	−6.9	−7.1
	Thiophenol	Chloroform	−2.6	−4.3	−6.9	−1.6	−5.2	−6.9	−7.6
19	Diethyl sulfide	Chloroform	−1.1	−4.0	−5.1	−1.2	−4.1	−5.3	−6.4
	Thioanisole	Carbon tetrachloride	−0.5	−5.2	−5.8	−0.6	−5.2	−5.8	−5.7
21	2-Propen-1-ol	Chloroform	−3.1	−0.3	−3.5	−2.8	−0.9	−3.7	−4.3
	m-Hydroxybenzaldehyde	1,2-Dichloroethane	−8.7	−2.0	−10.7	−7.7	−2.7	−10.4	−10.1
22	Fluorobenzene	Chloroform	−2.3	−1.6	−4.0	−1.7	−2.6	−4.3	−4.3
	Fluorobenzene	Carbon tetrachloride	−1.5	−1.9	−3.5	−1.1	−2.6	−3.7	−3.6
23	Trichloromethane	1-Iodohexadecane	−0.7	−2.9	−3.6	−0.8	−2.8	−3.7	−3.4
	p-Dichlorobenzene	Chloroform	−1.6	−4.3	−5.9	−1.0	−5.1	−6.1	−6.3
24	Bromobenzene	Chloroform	−2.1	−3.9	−6.1	−1.4	−4.9	−6.3	−6.1
	Bromobenzene	Carbon tetrachloride	−1.4	−4.4	−5.8	−0.9	−5.1	−6.0	−5.9
25	Iodobenzene	Chloroform	−2.0	−4.7	−6.7	−1.2	−5.6	−6.8	−6.6
	Iodobenzene	Carbon tetrachloride	−1.3	−5.2	−6.5	−0.8	−5.8	−6.6	−6.5
26	2,2,2-Trifluoroethanol	Chloroform	−6.4	1.5	−4.9	−6.1	1.4	−4.7	−3.0
	p-Bromophenol	Methylene chloride	−5.9	−3.9	−9.8	−4.7	−4.6	−9.3	−9.1
27	Ammonia	Carbon tetrachloride	−2.0	0.2	−1.8	−1.1	0.6	−0.5	−1.1
	Water	Chloroform	−6.4	2.8	−3.6	−6.0	2.7	−3.3	−2.1

Table 23. Selected data and results (kcal) in haloaromatic solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	$\Delta G_{\mathrm{S}}^{\circ}$
1	*n*-Octane	Fluorobenzene	−0.1	−4.7	−4.8	0.0	−4.7	−4.7	−5.0
	n-Octane	Bromobenzene	−0.1	−4.5	−4.6	0.0	−4.5	−4.5	−5.0
6	Toluene	Chlorobenzene	−2.5	−2.9	−5.4	−1.6	−3.7	−5.2	−5.2
	Toluene	Iodobenzene	−2.4	−2.8	−5.2	−1.5	−3.7	−5.2	−5.0
7	1-Hexanol	*o*-Dichlorobenzene	−3.3	−2.9	−6.2	−3.3	−3.2	−6.5	−5.7
	1-Heptanol	Chlorobenzene	−3.0	−3.9	−6.9	−3.0	−4.1	−7.1	−6.8
8	1,4-Dioxane	Bromobenzene	−3.2	−1.6	−4.8	−2.4	−2.3	−4.7	−5.0
	1,4-Dioxane	Iodobenzene	−3.1	−1.4	−4.5	−2.3	−2.3	−4.5	−4.9
10	Butanone	Chlorobenzene	−3.6	−1.1	−4.7	−3.6	−1.2	−4.8	−4.5
	2-Hexanone	Perfluorobenzene	−1.8	−3.0	−4.8	−1.9	−2.9	−4.8	−5.6
11	Propanoic acid	Chlorobenzene	−6.1	0.8	−5.3	−5.9	0.5	−5.4	−4.4
12	Propyl ethanoate	Bromobenzene	−4.3	−1.8	−6.1	−4.0	−2.1	−6.0	−4.9
	Butyl ethanoate	Perfluorobenzene	−2.4	−3.0	−5.4	−2.2	−3.1	−5.3	−5.5
13	Ethylamine	Iodobenzene	−1.5	−1.4	−2.9	−1.1	−1.6	−2.7	−2.7
	Propylamine	*o*-Dichlorobenzene	−1.7	−2.1	−3.8	−1.3	−2.3	−3.6	−3.4
14	Aniline	Chlorobenzene	−3.9	−3.4	−7.2	−3.4	−3.9	−7.4	−6.7
	Aniline	Bromobenzene	−3.8	−3.3	−7.2	−3.4	−4.0	−7.4	−6.7
26	*p*-Bromophenol	Chlorobenzene	−5.4	−4.0	−9.4	−4.4	−4.8	−9.1	−8.5
	p-Bromophenol	Iodobenzene	−5.1	−4.0	−9.1	−4.1	−4.9	−9.0	−8.5
27	Ammonia	Chlorobenzene	−3.1	0.3	−2.9	−1.7	0.7	−1.0	−1.2

Table 24. Selected data and results (kcal) in miscellaneous acidic solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	$\Delta G_{\mathrm{S}}^{\circ}$
1	*n*-Octane	*N*-Methylformamide	−0.1	−3.4	−3.5	0.0	−3.6	−3.6	−3.3
	n-Octane	Acetic acid	−0.1	−4.0	−4.0	0.0	−4.0	−4.1	−3.9
6	Toluene	Acetic acid	−2.6	−1.2	−3.8	−1.6	−2.4	−4.0	−4.5
	Toluene	2-Methoxyethanol	−3.0	−1.4	−4.4	−1.9	−2.7	−4.6	−4.5
7	Ethanol	2-Methoxyethanol	−4.3	−0.7	−5.0	−4.0	−1.1	−5.1	−4.7
	Ethanol	*N*-Methylformamide	−4.6	−0.1	−4.7	−4.3	−0.6	−4.8	−5.1
8	1,4-Dioxane	*N*-Methylformamide	−4.0	−1.3	−5.3	−3.0	−2.2	−5.2	−4.9
	1,4-Dioxane	Acetic acid	−3.3	−2.0	−5.3	−2.4	−2.8	−5.3	−5.8
10	Butanone	2-Methoxyethanol	−4.3	−0.5	−4.9	−4.4	−0.5	−4.8	−4.3
	Butanone	*N*-Methylformamide	−4.7	−0.2	−4.9	−4.7	−0.2	−5.0	−4.3

Table 25. Selected data and results (kcal) in miscellaneous non-acidic solvents

Solute class	Solute	Solvent	SM5.4/AM1			SM5.4/PM3			Experiment
			ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	ΔG_{ENP}	G_{CDS}	$\Delta G_{\mathrm{S}}^{\circ}$	$\Delta G_{\mathrm{S}}^{\circ}$
1	*n*-Octane	Sulfolane	−0.1	−2.6	−2.7	0.0	−2.9	−2.9	−2.4
	n-Octane	Carbon disulfide	−0.1	−5.1	−5.2	0.0	−5.1	−5.1	−5.7
6	Toluene	Dimethylsulfoxide	−3.1	−1.3	−4.4	−2.0	−2.6	−4.5	−4.4
	Toluene	Carbon disulfide	−1.8	−3.3	−5.2	−1.1	−4.1	−5.3	−5.4
7	Phenol	Carbon disulfide	−4.0	−2.5	−6.5	−3.1	−3.3	−6.5	−6.3
	1-Heptanol	Tributyl phosphate	−3.2	−4.8	−7.9	−3.2	−5.0	−8.2	−8.0
8	1,4-Dioxane	Sulfolane	−4.0	−0.3	−4.3	−2.9	−1.2	−4.2	−4.9
	1,4-Dioxane	Dimethylsulfoxide	−4.0	−0.8	−4.8	−2.9	−1.6	−4.5	−4.9
10	Propanone	Carbon disulfide	−2.7	−0.5	−3.2	−2.7	−0.5	−3.2	−3.1
	Butanone	Sulfolane	−4.6	0.5	−4.1	−4.6	0.4	−4.2	−4.1
11	Ethanoic acid	Tributyl phosphate	−7.1	−0.1	−7.3	−7.0	−0.4	−7.3	−7.1
	Butanoic acid	Tributyl phosphate	−6.3	−1.8	−8.1	−6.1	−2.0	−8.1	−8.3
12	Methyl ethanoate	Carbon disulfide	−3.3	−0.1	−3.4	−3.0	−0.5	−3.4	−3.7
	Ethyl ethanoate	Carbon disulfide	−3.2	−1.4	−4.5	−2.9	−1.6	−4.5	−4.1
13	Ethylamine	Tributyl phosphate	−1.8	−1.7	−3.5	−1.2	−2.0	−3.3	−3.3
	Butylamine	Tributyl phosphate	−1.6	−3.2	−4.8	−1.2	−3.4	−4.6	−4.3
14	Aniline	Tributyl phosphate	−4.1	−3.1	−7.3	−3.7	−4.4	−8.0	−7.6
16	1-Nitropropane	Carbon disulfide	−4.8	−0.4	−5.2	−4.1	−0.7	−4.7	−4.5
21	2-Methoxyethanol	Tributyl phosphate	−4.8	−2.1	−6.9	−4.4	−2.6	−7.0	−6.1
27	Water	Tributyl phosphate	−7.2	−0.6	−7.8	−6.8	−0.8	−7.6	−4.7

Table 26. Experimental solvation free energies (kcal) and solvent properties in water and organic solvents

Solvent	Solutes					n	α	β	γ	ε
	n-octane	toluene	1,4-dioxane	butanone	ethanol					
n-Hexadecane	−5.0	−4.5	−3.8	−3.1	−2.0	1.43	0.00	0.00	38.9	2.06
1-Octanol	−4.2	−4.6	−4.9	−3.8	−4.4	1.43	0.37	0.48	39.0	9.87
Formamide		−2.9	−4.8	−4.0	−4.8	1.45	0.62	0.60	82.1	109.
Acetic acid	−3.9	−4.5	−5.8	−4.8	−5.3	1.37	0.61	0.44	39.0	6.25
Ethanol	−4.2	−4.6	−4.7	−4.3		1.36	0.37	0.48	31.6	24.9
1,2-Ethanediol		−2.8	−4.1	−3.3	−4.7	1.43	0.58	0.78	69.0	40.3
Methanol	−3.8	−4.3	−4.9	−4.6		1.33	0.43	0.47	22.1	32.6
Water	2.9	−0.9	−5.1	−3.6	−5.0	1.33	0.82	0.35	104.0	78.3

Table 27. Free energies of solvation (kcal) for four solutes in the solvents 1,2-ethanediol and formamide

Solute	1,2-Ethanediol			Formamide		
	SM5.4/AM1	SM5.4/PM3	Expt.	SM5.4/AM1	SM5.4/PM3	Expt.
Toluene	−3.0	−3.4	−2.8	−2.6	−3.0	−2.9
Ethanol	−4.5	−4.7	−4.7	−4.0	−4.3	−4.8
1,4-Dioxane	−4.8	−4.9	−4.1	−4.5	−4.6	−4.8
Butanone	−4.0	−4.1	−3.3	−3.7	−3.9	−4.0

Table 28. Solvent properties and free energies of solvation (kcal) for ethanol in a variety of solvents

Solvent	n	α	β	γ	ε	$1 - (1/\varepsilon)$	ΔG_S°	
							SM5.4/AM1	Experimental
N,*N*-Dimethylacetamide	1.44	0.00	0.78	47.7	37.8	0.97	−4.9	−5.4
DMSO	1.42	0.00	0.88	61.8	46.8	0.98	−4.5	−5.3
Acetic acid	1.37	0.61	0.44	39.0	6.3	0.84	−4.4	−5.3
N-Methylformamide	1.43	0.40	0.55	55.7	181.6	0.99	−4.7	−5.1
1-Butanol	1.40	0.37	0.48	35.9	17.3	0.94	−5.0	−5.0
Tributylphosphate	1.42	0.00	1.21	37.4	8.2	0.88	−5.1	−4.6
Tetrahydrofuran	1.41	0.00	0.48	38.0	7.4	0.87	−4.3	−4.6
1-Octanol	1.43	0.37	0.48	39.0	9.9	0.90	−4.7	−4.4
Acetonitrile	1.34	0.07	0.32	41.3	37.5	0.98	−4.6	−4.4
Diethyl ether	1.35	0.00	0.41	24.0	4.2	0.76	−4.1	−4.4
Sulfolane	1.48	0.00	0.88	77.6	44.0	0.98	−4.0	−4.3
Acetophenone	1.54	0.00	0.48	56.2	17.4	0.94	−4.2	−4.1
Chloroform	1.45	0.15	0.02	38.4	4.7	0.79	−3.6	−3.9
Triethylamine	1.40	0.00	0.79	29.1	2.4	0.58	−3.5	−3.9
Dibutyl ether	1.40	0.00	0.45	32.3	3.0	0.67	−3.5	−3.5
Toluene	1.50	0.00	0.14	40.2	2.4	0.58	−3.0	−3.3
Carbon disulfide	1.63	0.00	0.07	45.5	2.6	0.62	−2.6	−3.0
CCl_4	1.46	0.00	0.00	38.0	2.2	0.55	−2.3	−3.0
n-Penatane	1.36	0.00	0.00	22.3	1.8	0.46	−2.3	−2.2
n-Hexadecane	1.43	0.00	0.00	38.9	2.1	0.51	−2.1	−2.0
R^{2}[a]	0.06	0.21	0.48	0.13	0.19	0.77	0.86	

[a] Correlation of solvent property or calculated free energy of solvation with the experimental free energy of solvation for solvents in this table

ute a considerably larger fraction of the total solvation free energy to solute/solvent polarization. This difference can be attributed to the different solute cavities employed – Luque et al. employed standard van der Waals radii [60] scaled by a factor of 1.8, making for a much larger cavity than that formed from the SM5.4 coulomb radii [19]. Reducing the magnitude of the solute/solvent polarization in the self-consistent reaction field calculations causes the atomic surface tensions to adjust to recapture this energy during the parameterization. Indeed, we have shown that it is possible to develop accurate solvation models that employ *only* atomic surface tensions (i.e., ΔG_{ENP} is set to zero) [38]. However, when a physical partitioning of the electrostatic and non-electrostatic components of solvation in carbon tetrachloride is desirable, it appears that the SM5.4 models should be preferred over those of Luque et al.

6 Application of the SM5.4 parameterizations

6.1 Partition coefficients

The capability of the SM5.4 models to calculate partition coefficients between two organic solvents is

Table 29. SM5.4, MST/ST, and experimental free energies of solvation (kcal/mol) in CCl_4

Solute	SM5.4		MST/ST			Expt.
	AM1	PM3	AM1	PM3	6-31G*	
Methanol	−1.4	−1.4	−2.5	−2.5	−2.6	−2.3
Ethanol	−2.3	−2.3	−3.0	−2.9	−3.0	−3.0
n-Propanol	−3.0	−3.0	−3.5	−3.5	−3.6	−3.6
n-Butanol	−3.6	−3.8	−4.1	−4.1	−4.2	−4.2
n-Pentanol	−4.4	−4.5	−4.7	−4.7	−4.8	−4.7
n-Hexanol	−5.2	−5.4	−5.3	−5.3	−5.4	−5.0
n-Heptanol	−6.0	−6.1	−5.9	−5.9	−6.1	−6.5
Phenol	−5.8	−5.8	−5.8	−5.7	−5.9	−6.1
o-Cresol	−6.4	−6.2	−6.2	−6.1	−6.2	−6.5
p-Cresol	−6.5	−6.3	−6.4	−6.2	−6.5	−6.3
Propanone	−2.9	−3.0	−4.0	−4.1	−4.2	−3.4
Acetophenone	−6.6	−6.8	−7.1	−7.1	−7.1	−7.1
Ethanoic acid	−2.6	−2.8	−4.3	−4.2	−4.2	−3.6
Propanoic acid	−3.3	−3.4	−4.6	−4.5	−4.6	−4.1
Butanoic acid	−3.9	−4.0	−5.2	−5.2	−5.1	−4.8
Methyl acetate	−3.1	−3.1	−4.7	−4.6	−4.5	−3.8
Ethyl acetate	−4.1	−4.1	−5.2	−5.1	−5.0	−4.4
Ethylamine	−2.7	−2.6	−3.0	−3.0	−3.0	−2.8
n-Propylamine	−3.4	−3.3	−3.5	−3.6	−3.6	−3.6
n-Butylamine	−4.1	−4.0	−4.1	−4.2	−4.2	−4.3
Aniline	−6.0	−6.2	−5.7	−5.7	−5.9	−6.1
Diethylamine	−3.5	−3.7	−3.9	−4.1	−3.7	−4.1
Trimethylamine	−3.0	−2.9	−3.2	−3.3	−3.1	−3.1
Pyridine	−4.8	−4.9	−4.9	−5.0	−5.0	−5.0
m-Nitrophenol	−8.0	−8.1	−8.5	−8.7	−8.4	−8.8
p-Nitrophenol	−8.3	−8.4	−8.5	−8.7	−8.5	−9.4
p-Hydroxybenzaldehyde	−7.6	−7.7	−7.7	−7.6	−7.9	−8.2
Mean unsigned error	0.5	0.4	0.3	0.3	0.3	

examined in Table 30. Just as we calculated free energies from partition coefficients using Eqs. (1)–(3), we can also use free energies of solvation to calculate partition coefficients. We can calculate over 20 000 partition coefficients with the 1786 free energies of solvation used to develop the parameterizations presented here. Table 30 gives the errors in the predictions of the SM5.4/AM1 and SM5.4/PM3 parameterizations for these organic/organic partition coefficients. Using the SM5.4-aqueous parameterizations, we can now calculate octanol/water (or any organic/water) partition coefficients. The mean unsigned error in $\log_{10} P_{A/B}$ values is 0.12 $\log_{10}$ units, corresponding to an error of a factor of 1.3 in the partitioning equilibrium constant.

Table 30. Calculation of organic/organic partition coeficients ($\log_{10} P_{A/B}$)

	SM5.4/A	SM5.4/P
Total number	20 922	20 916[a]
Mean signed error	−0.01	−0.01
Mean unsigned error	0.12	0.12

[a] Number differs due to the removal of hydrazine from the SM5.4/PM3 data set

Table 31. Free energy change (kcal) for *trans* → *gauche* conformational transition in 1,2-dichloroethane

Solvent	SM5.4/AM1	Experiment
Tetrachloroethene	−0.37	−0.31
Carbon disulfide	−0.42	−0.36
Ethyl ether	−0.64	−0.50
Ethyl acetate	−0.74	−0.76
Dichloroethane	−0.85	−0.88
Acetone	−0.93	−1.00
Acetonitrile	−1.01	−1.03

6.2 Conformational equilibria

Table 31 gives an example application of the SM5.4/AM1 parameterization to a conformational problem. In 1,2-dichloroethane, the chlorine atoms can be either *gauche* or *trans* to each other. The difference in the free energies of solvation of the two isomers has been measured experimentally [85] in a number of solvents, with the *gauche* isomer being increasingly stabilized versus the *trans* in the more polar solvents. Here we see that results calculated by SM5.4/AM1 agree well with the experimental measurements [21].

6.3 Nucleic acids

We have applied the aqueous and chloroform parameterizations to predict solvation energies and partition coefficients of natural and unnatural nucleic acid bases; this work is described elsewhere [20].

7 Concluding remarks

We have presented two new quantum mechanical parameterizations for solvation free energies in any organic solvent for which five widely available macroscopic properties of the solvent are known. The model includes accurate atomic partial charges, long-range electrostatic effects, medium-range solvent structural effects, and short-range effects due to dispersion and hydrogen bonding. The parameterizations are based on the AM1 and PM3 semi-empirical Hamiltonians, which results in fast and accurate calculation of the free energy of solvation. The required solvent properties are the dielectric constant [75], the index of refraction [75], the macroscopic surface tension [75, 86], and Abraham's $\sum \alpha_2^H$ and $\sum \beta_2^H$ descriptors [74, 78]. The parameterizations were developed using experimental free energies of solvation involving 1786 solute/solvent pairs, and they have uniformly low mean errors over the wide range of solutes and solvents considered. The chief result of the present paper is that by adding the large solvent radius for calculating intermediate-range effects [10, 11] and by parameterizing the surface tension coefficients in terms of only a few macroscopic solvent descriptors, we were able to extend the SM5.4/AM1-aqueous model with comparable accuracy to essentially all organic solvents. There are very few published experimental data for solvation energies in solvents that are miscible with water, but the present model should be as applicable to such solvents as to those for which data do exist, so it can fill a significant gap in the database. The parameterizations presented here can also be combined with the recently developed SM5.4-aqueous [19] and SM5.4-chloroform [24] parameterizations to calculate virtually any partition coefficient of interest. The SM5.4-aqueous, -organic, -chloroform, -benzene, and -toluene parameterizations are available in AMSOL, beginning with version 6.0 [87].

Acknowledgements. D.J.G. thanks the Eastman-Kodak Company for a fellowship. We are grateful to Zhen (Michael) Gu, and Candee Chambers for assistance. This work was supported in part by the National Science Foundation under grants no. CHE94-23927 and CHE95-25819. Some of the material in this paper was reproduced with permission from a Ph.D thesis [88].

References

1. Reichardt C (1990) Solvents and solvent effects in organic chemistry. VCH, New York
2. Honig B, Sharp K, Yang A-S (1993) J Phys Chem 97: 1101
3. Tomasi J, Persico M (1994) Chem Rev 94: 2027
4. Cramer CJ, Truhlar DG (eds) (1994) Structure and reactivity in aqueous solution. American Chemical Society, Washington, DC
5. Cramer CJ, Truhlar DG (1995) In: Lipkowitz KB, Boyd DB (eds) Reviews in computational chemistry. VCH, New York, p 1
6. Cramer CJ, Truhlar DG (1996) In: Tapia O, Bertrán J (eds) Solvent effects and chemical reactivity. Kluwer, Dordrecht, p 1
7. Bonaccorsi R, Floris F, Tomasi J (1990) J Mol Liquids 47: 25
8. Bonaccorsi R, Floris F, Palla P, Tomasi J (1990) Theor Chim Acta 162: 213
9. Storer JW, Giesen DJ, Hawkins GD, Lynch GC, Cramer CJ, Truhlar DG, Liotard DA (1994) In: Cramer CJ, Truhlar DG (eds) Structure and reactivity in aqueous solution. American Chemical Society, Washington, DC, p 24
10. Giesen DJ, Storer JW, Cramer CJ, Truhlar DG (1995) J Am Chem Soc 117: 1057
11. Giesen DJ, Cramer CJ, Truhlar DG (1995) J Phys Chem 9: 7137
12. Sitkoff D, Ben-Tal N, Honig B (1996) J Phys Chem 100: 2744
13. Luque FJ, Zhang Y, Alemán C, Bachs M, Gao J, Orozco M (1996) J Phys Chem 100: 4269
14. Luque FJ, Alemán C, Bachs M, Orozco M (1996) J Comp Chem 17: 806
15. Leo AJ (1993) Chem Rev 93: 1281
16. Abraham MH, Whiting GS, Fuchs R, Chambers EJ (1990) J Chem Soc, Perkin Trans 2: 291
17. Zhang Y, Dallas AJ, Carr PW (1993) J Chrom 638: 43
18. Leo AJ (1994) Masterfile database from MedChem Software
19. Chambers CC, Hawkins GD, Cramer CJ, Truhlar DG (1996) J Phys Chem 100: 16385
20. Giesen DJ, Chambers CC, Cramer CJ, Truhlar DG (1997) J Phys Chem 101: 5084
21. Giesen DJ, Chambers CC, Hawkins GD, Cramer CJ, Truhlar DG In: Irikura K, Frurip DJ (eds) Computational thermochemistry: prediction and estimation of molecular thermodynamics. American Chemical Society, Washington, DC (in press)
22. Leo A, Hansch C, Elkins D (1971) Chem Rev 71: 525
23. Giesen DJ, Gu MZ, Cramer CJ, Truhlar DG (1996) J Org Chem 61: 8720
24. Giesen DJ, Chambers CC, Cramer CJ, Truhlar DG (1997) J Phys Chem 101: 2061
25. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107: 3902
26. Dewar MJS, Zoebisch EG (1988) J Mol Struct (Theochem) 180: 1
27. Dewar MJS, Yate-Ching Y (1990) Inorg Chem 29: 3881
28. Stewart JJP (1989) J Comp Chem 10: 221
29. Dallas AJ (1995) Ph.D. thesis, University of Minnesota, Minneapolis
30. Kohata K, Fukuyama T, Kuchitu K (1982) J Phys Chem 86: 602
31. Cramer CJ, Truhlar DG (1991) J Am Chem Soc 113: 8305
32. Cramer CJ, Truhlar DG (1992) Science 256: 213
33. Cramer CJ, Truhlar DG (1992) J Comp Chem 13: 1089
34. Hawkins GD, Cramer CJ, Truhlar DG (1995) Chem Phys Lett 246: 122
35. Liotard DA, Hawkins GD, Lynch GC, Cramer CJ, Truhlar DG (1995) J Comp Chem 16: 422
36. Barrows SE, Dulles FJ, Cramer CJ, Truhlar DG, French AD (1995) Carbohydr Res 276: 219
37. Hawkins GD, Cramer CJ, Truhlar DG (1996) J Phys Chem 100: 19824
38. Hawkins GD, Cramer CJ, Truhlar DG J Phys Chem (in press)
39. Storer JW, Giesen DJ, Cramer CJ, Truhlar DG (1995) J Comput Aided Mol Des 9: 87
40. Thiel W, Voityuk AA (1992) Int J Quantum Chem 44: 807
41. Thiel W, Voityuk AA (1992) Theor Chim Acta 81: 391
42. Thiel W, Voityuk AA (1996) J Phys Chem 100: 616
43. Becke AD (1988) Phys Rev A 38: 3098
44. Perdew J, Wang Y (1992) Phys Rev B 45: 13244
45. Cramer CJ, Truhlar DG (1992) J Comput Aided Mol Des 6: 629
46. Nicholls A, Honig B (1991) J Comp Chem 12: 435
47. Sharp K (1991) J Comp Chem 12: 454
48. Gilson MK, Davis ME, Luty BA, McCammon JA (1993) J Phys Chem 97: 3591
49. Smith PE, Pettitt BM (1994) J Phys Chem 98: 9700
50. Chen Y, Noodleman L, Case DA, Bashford D (1994) J Phys Chem 98: 11059
51. Schaefer M, Karplus M (1996) J Phys Chem 100: 1578
52. Hoijtink GJ, de Boer E, Van der Meij PH, Weijland WP (1956) Recl Trav Chim Pays-Bas 75: 487
53. Peradejordi F (1963) Cah Phys 17: 343
54. Jano I (1965) C R Acad Sci Paris 261: 103
55. Tapia O (1980) In: Daudel R, Pullman A, Salem L, Viellard A (eds) Quantum theory of chemical reactions. Reidel, Dordrecht, p 25
56. Tucker SC, Truhlar DG (1989) Chem Phys Lett 157: 164
57. Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) J Am Chem Soc 112: 6127
58. Kikuchi O, Matsuoka T, Sawahata H, Takahashi O (1994) J Mol Struct (Theochem) 305: 79
59. Lee B, Richards FM (1971) J Mol Biol 55: 379
60. Bondi A (1964) J Phys Chem 68: 441
61. Fowler FW, Katritzky AR, Rutherford RJD (1971) J Chem Soc B 460
62. Koppel IA, Palm VA (1972) In: Chapman NB, Shorter J (eds) Advances in linear free energy relationships. Plenum Press, London, p 203
63. Fawcett WR, Krygowski TM (1975) Aust J Chem 28: 2115
64. Krygowski TM, Fawcett WR (1975) J Am Chem Soc 97: 2143
65. Fawcett WR, Krygowski TM (1976) Can J Chem 54: 3283
66. Kamlet MJ, Abboud JL, Taft RW (1977) J Am Chem Soc 99: 6027
67. Jackson WG, Lawrence GA, Lay PA, Sargeson AM (1982) Aust J Chem 35: 1561
68. Svoboda P, Pytela O, Vecera M (1983) Collect Czech Chem Commun 48: 3287
69. Glikberg S, Marcus Y (1983) J Solution Chem 12: 255
70. Marcus Y (1991) J Phys Chem 95: 8886
71. Streck C, Richert R (1994) Ber Bunsen-Ges Phys Chem 98: 619
72. Laurence C, Nicolet P, Dalati MT, Abboud J-LM, Notario R (1994) J Phys Chem 98: 5807
73. Gutmann V, Resch G (1995) Lecture notes on solution chemistry. World Scientific, Singapore
74. Abraham MH (1993) Chem Soc Rev 73

75. Lide DR (ed) (1995) CRC handbook of chemistry and physics, 75th edn. CRC Press, Boca Raton
76. Barton AFM (1991) CRC handbook of solubility parameters and other cohesion parameters, 2n edn. CRC Press, Boca Raton
77. Abraham MH, Grellier PL, Prior DV, Duce PP, Morris JJ, Taylor P (1989) J Chem Soc, Perkin Trans 2: 699
78. Abraham MH (1994) In: Politzer P, Murray JS (eds) Quantitative treatments of solute/solvent interactions. Elsevier, Amsterdam, p 83
79. Abraham MH (1993) J Phys Org Chem 6: 660
80. Murray JS, Brinck T, Lane P, Paulsen K, Politzer P (1994) J Mol Struct (Theochem) 307: 55
81. Grunwald E, Pan K-C (1976) J Phys Chem 80: 2929
82. Grunwald E, Anderson SP, Effio A, Gould SE, Pan K-C (1976) J Phys Chem 80: 2935
83. Grunwald E, Pan K-C, Effio A (1976) J Phys Chem 80: 2937
84. Miertus S, Scrocco E, Tomasi J (1981) Chem Phys 55: 117
85. Depaepe J-M, Ryckaert J-P (1995) Chem Phys Lett 245: 653
86. Jasper JJ (1972) J Phys Chem Ref Data 1: 841
87 Hawkins GD, Giesen DJ, Lynch GC, Chambers CC, Rossi I, Storer JW, Rinaldi D, Liotard DA, Cramer CJ, Truhlar DG (1997) AMSOL-version 6.1.1, Oxford Molecular Group, London
88. Giesen DJ (1997) Ph.D. thesis, University of Minnesota