

# Extension of a Temperature-Dependent Aqueous Solvation Model to Compounds Containing Nitrogen, Fluorine, Chlorine, Bromine, and Sulfur

Adam C. Chamberlin, Christopher J. Cramer,\* and Donald G. Truhlar\*

Department of Chemistry and Supercomputing Institute, 207 Pleasant Street SE, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received: August 20, 2007; In Final Form: December 3, 2007

Most methods for predicting free energies of solvation have been developed or validated exclusively for room temperature. Recently, we developed a model called SM6T for predicting aqueous solvation free energies as a function of temperature for solutes composed of C, H, or O, and here we present solvation model 8 with temperature dependence (SM8T) for predicting the temperature dependence of aqueous free energies of solvation for compounds containing H, C, N, O, F, S, Cl, and Br in the range 273–373 K. We also describe the database of experimental aqueous free energies of solvation used to parametrize the model. SM8T partitions the temperature dependence of the free energy of solvation into two components: the temperature dependence of the bulk electrostatic contribution to the free energy of solvation, which is computed using the generalized Born equation, and the temperature dependence of first-solvation-shell effects, which is modeled by terms proportional to the solvent-exposed surface areas of atoms in functional groups determined entirely by geometry. SM8T predicts the temperature dependence of aqueous free energies of solvation with a mean unsigned error of 0.08 kcal/mol over a database of 4403 measurements on 348 compounds at various temperatures. We also discuss the accuracy of SM8T for predicting the temperature dependence of aqueous free energies of solvation for ions and present free energies of solvation as a function of temperature for two sample ions.

## 1. Introduction

Free energies of solvation play a key role in the computation of solubilities, vapor pressures, and partition coefficients.<sup>1–4</sup> Implicit solvation models<sup>5–8</sup> offer a particularly efficient way to predict such free energies of solvation, and they have been widely applied to such diverse areas as drug design,<sup>3,9–14</sup> biomolecular property prediction,<sup>10,15–21</sup> atmospheric and soil pollutant partitioning,<sup>22–24</sup> and prediction of the properties of energetic molecules.<sup>25</sup> However, the effect of temperature upon the free energy of solvation (and solubility) has been largely ignored in the development of continuum solvation models, which have been developed or validated almost exclusively to model solvation at 298 K. This is despite the fact that solubilities may vary by more than 2 orders of magnitude over the temperature range of 273–373 K.

In a prior paper,<sup>26</sup> we introduced an extension of our SM6<sup>27</sup> aqueous solvation model, SM6T, that accounted for the temperature dependence of free energies of solvation for compounds containing H, C, or O. The objective of the present paper is to extend our ability to predict the temperature dependence of aqueous free energies of solvation to compounds containing N, F, Cl, Br, and S in aqueous solution over the temperature range 273–373 K. Since the time when SM6T was created, we have developed a new universal solvation model called SM8<sup>28</sup> in which “universal” refers to the fact that it is applicable to all organic solvents as well as to water. The aqueous version of SM8 is almost identical to the aqueous SM6 model, and it has the same Coulomb radii for aqueous solution. Thus, the SM6T parameters work equally well with SM8 as with SM6. Therefore,

the extension presented here is made using SM8 rather than SM6. We refer to the new model as SM8T (solvation model 8 with temperature dependence).

The temperature dependence of the standard-state free energy of solvation,  $\Delta G_s^*$  is given by<sup>2,29,30</sup>

$$\Delta G_s^*(T) = \Delta G_s^*(T_0) - \Delta S_s^*(T_0)[T - T_0] + \Delta C_{p,s}^*(T_0)[(T - T_0) - T \ln(T/T_0)] \quad (1)$$

where  $T$  is the temperature,  $T_0$  is a reference temperature taken throughout as 298 K, and  $\Delta S_s^*$  and  $\Delta C_{p,s}^*$  are, respectively, the standard-state entropy and heat capacity of solvation. Equation 1 is true for any standard-state choice; the choice denoted by the \* superscript is explained in section 2.A. Ignoring the temperature dependence of the enthalpies,  $\Delta H_s^*$ , and entropies,  $\Delta S_s^*$ , of solvation yields a linear dependence of  $\Delta G_s^*$  on  $T$ . Heat capacities of solvation and, hence, temperature-dependent enthalpies and entropies of solvation have been shown,<sup>26,31,32</sup> however, to be significant over the 273–373 K temperature range for many compounds. Experiment has shown that  $\Delta C_{p,s}^*$  can itself depend on temperature, but we take it to be a constant in SM8T.

In Section 2, we discuss the database of experimental data for compounds containing N, F, Cl, Br, and S. In Section 3, we extend SM6T to predict the temperature dependence of the free energy of solvation for these new compounds. In Section 4, we will discuss the current status of SM8T and our plans for further development.

## 2. Developing the Database

**2.A. Literature Search.** We obtained experimental free energies of solvation using a variety of experimental data,

\* To whom correspondence should be addressed. E-mails: cramer@chem.umn.edu, truhlar@umn.edu.

including Henry's Law constants,<sup>33</sup> solubilities<sup>34,35</sup> in conjunction with vapor pressures,<sup>36,37</sup> activity coefficients<sup>38–40</sup> in conjunction with vapor pressures, and free energies of solvation at single temperatures<sup>31,41,42</sup> in conjunction with enthalpies<sup>31,42,43</sup> and heat capacities of solvation.<sup>31</sup> The free energy of solvation corresponds to the equilibrium of a solute (A) between the gas phase (g) and solution; we focus exclusively here on water (aq) as the solvent, so the solvation process becomes



All the free energies of solvation reported here correspond to a 1 M ideal gas and a 1 M ideal infinitely dilute solution; such free energies of solvation are denoted as  $\Delta G_S^*$ . This convention eliminates any change in the translational entropy upon solvation, as discussed by Ben-Naim.<sup>44</sup> However, experimentally determined standard-state free energies of solvation are usually reported for a 1 bar ideal gas and a 1 M ideal solution. These choices yield a quantity called  $\Delta G_S^\circ$ . To convert the free energies of solvation in the present article to a gas-phase standard state of 1 bar,  $\Delta G_S^\circ$ , one must use the following equation,

$$\Delta G_S^\circ = \Delta G_S^* + RT \ln(RTC^*/P^\circ) \quad (3)$$

where  $R$  is the gas constant with the appropriate units ( $8.314 \text{ J K}^{-1} \text{ mol}^{-1}$  or  $8.314 \times 10^{-2} \text{ L bar K}^{-1} \text{ mol}^{-1}$ ),  $C^*$  is 1 M, and  $P^\circ$  is 1 bar.

In the case of an infinitely dilute solution, the partition coefficient for this process corresponds to the Henry's Law coefficient,  $K_H^*$ , in which the standard-state free energy of solvation for this process may be computed as

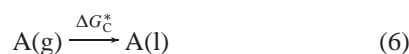
$$\Delta G_S^* = RT \ln(K_H^*/K^*) \quad (4)$$

where  $K^*$  is unity in the units of the standard states in the gas phase and solution. In this case,  $K^* = 1 \text{ M gas phase}/1 \text{ M solution}$ , and  $T$  is the temperature at which the measurement was taken. Henry's Law constants are tabulated using a variety of units, each of which corresponds to a different standard state, and so care was taken to ensure that all reported Henry's Law constants used in this work were converted to

$$K_H^* = \frac{\text{concentration (M) in gas phase}}{\text{concentration (M) in liquid solution}} \quad (5)$$

Staudinger and Roberts<sup>33</sup> provide further discussion of this particular issue.

Another approach to obtaining experimental free energies of solvation is to use solubilities and the vapor pressure of the pure solute. This approach results from treating the process of solvation as a series of steps. In the first step, the solute, A, condenses from the gas phase into its pure liquid state (l)



The standard-state free energy associated with moving a solute molecule from a 1 M ideal gas to a 1 M ideal pure liquid composed of the solute may be computed as<sup>4,45</sup>

$$\Delta G_C^* = RT \ln\left(\frac{P_A^*}{RTM_A^l}\right) \quad (7)$$

where  $P_A^*$  is the saturated vapor pressure of A over pure liquid A, and  $M_A^l$  is the molarity of the solute in its pure liquid state.

The second step of the process involves transferring the solute from its pure liquid phase to aqueous solution



The standard-state free energy associated with moving a solute molecule from a 1 M ideal neat liquid (composed of the solute) to a 1 M ideal aqueous solution may be computed as<sup>4</sup>

$$\Delta G_L^* = -RT \ln \frac{M_A^{aq}}{M_A^l} \quad (9)$$

where  $M_A^{aq}$  is the aqueous molar concentration of A in equilibrium with neat liquid at temperature  $T$ . The combination of eq 7 and 9 yields

$$\Delta G_S^* = RT \ln \frac{P_A^*}{RTM_A^{aq}} \quad (10)$$

The equilibria used to compute the standard-state free energies of solvation are for the case in which the solution is infinitely dilute, that is, for the case of no solute–solute interactions. However, we have found in prior work<sup>4</sup> that the assumption of an ideal solution when computing free energies of solvation using solubilities does not introduce errors that are significantly greater than the inherent accuracy of our solvation models.

The vapor pressures used in the development of this database were computed using empirical fits<sup>46,47</sup> to experimental data. The errors associated with these fits<sup>46,47</sup> are comparable to the experimental errors associated with measurement of solubilities<sup>34,48</sup> and Henry's Law constants.<sup>49–52</sup>

A third approach to computing standard-state free energies of solvation is to use a combination of activity coefficients at infinite dilution and vapor pressures. The activity coefficient is a unitless constant of proportionality between the Henry's Law constant for a given solute and the vapor pressure of the pure solute.<sup>2,29,30</sup> It is written as<sup>3,38,53</sup>

$$\gamma_A^\infty = \frac{P^\circ K_H^X}{P_A^*} \quad (11)$$

where  $K_H^X$  is the Henry's law constant in units of mole fraction in the gas and liquid phases, and  $P^\circ$  is the standard-state vapor pressure. The standard-state free energy of solvation is calculated from the activity coefficient as<sup>26</sup>

$$\Delta G_S^* = RT \ln\left(\frac{\gamma_A^\infty P_A^*}{RTM_w^*}\right) \quad (12)$$

where  $M_w^*$  is the molarity of pure water. For a derivation of the relation between the activity coefficient and the free energy of solvation, see Chamberlin et al.<sup>26</sup>

A final means to compute standard-state free energies of solvation is to use free energies of solvation at a single reference temperature in conjunction with enthalpies and heat capacities of solvation. There are a number of equations relating standard thermodynamic properties to free energies of solvation. One common representation popularized by Glew and Clark<sup>54</sup> is the free energy of solvation as a function of temperature, enthalpy, heat capacity, and the derivatives of heat capacity with respect to temperature,

$$\Delta G_S^* = \Delta G_S^*(T_0) \frac{T}{T_0} - \Delta H_S^*(T_0) \left[ \frac{T}{T_0} - 1 \right] - \Delta C_{p,S}^*(T_0) \left[ T_0 - T + T \ln \left( \frac{T}{T_0} \right) \right] - \left( \frac{d\Delta C_{p,S}^*(T_0)}{dT} \right)_{T_0} \dots \quad (13)$$

where  $d\Delta C_{p,S}^*(T_0)/dT$  is the temperature dependence of the heat capacity of solvation at the reference temperature  $T_0$ . In the interest of obtaining sufficient amounts of data, we disregarded the temperature dependence of the heat capacity of solvation, if it was not readily available, and we disregarded all of the higher terms, which resulted in the following equation:

$$\Delta G_S^* = \Delta G_S^*(T_0) \left[ \frac{T}{T_0} \right] + \Delta H_S^*(T_0) \left[ 1 - \frac{T}{T_0} \right] + \Delta C_{p,S}^*(T_0) \left[ (T - T_0) - T \ln \left( \frac{T}{T_0} \right) \right] \quad (14)$$

The reader should note that by substituting

$$\Delta G_S^*(T_0) = \Delta H_S^*(T_0) - T_0 \Delta S_S^*(T_0) \quad (15)$$

into eq 14, one regains eq 1.

**2.B. Data Analysis and Removal.** There is a large body of experimental data for neutral solutes. Once we had a sufficiently extensive collection of data for neutral solutes, we analyzed it and discarded questionable data by using the following criteria:

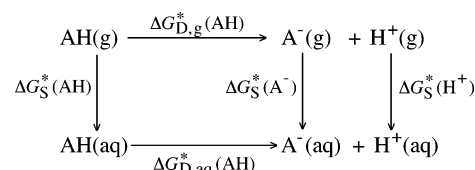
(1) We discarded data from any source that provided experimental data at only a single temperature for a given compound, except for some free energies of solvation at 298 K. We used this particular criterion because combining such data with data at other temperatures from other sources may lead to spurious temperature dependence. The reason an exception was made was because it is absolutely critical to have a data point at 298 K that can be used to estimate the temperature dependence of the remaining data. Some compounds did not have data that spanned 298 K, which could have been used to interpolate a value at 298 K. In the rare instances that we used a single point, the point was taken either from our SMx solvation database<sup>55</sup> or from some other reliable source.

(2) We eliminated all experimental data from sources where the derived experimental free energies of solvation systematically deviated by 0.4 kcal/mol or more from values obtained from two or more other sources.

(3) We discarded any compound for which less than three experimental points were available. To estimate the curvature of the temperature dependence, at least three data points are necessary. In most cases, we discarded experimental data if it had less than four points. However, in the interest of including some compounds with very unique functionality, such as chloropicrin, we retained a total of six compounds having only three experimental points in the new dataset.

(4) We required that no data point deviate by more than two standard deviations from a fit to all of the remaining experimental data for a given compound. Once a data point was removed, the fit was recalculated using the remaining data. This process was iterated until none of the remaining points deviated by more than two standard deviations from the fit. This criterion was applied to eliminate nonsystematic errors in the experimental data.

## SCHEME 1



(5) We did not include data from any source that we concluded was unreliable either because it consistently disagreed with other workers or because the authors did not provide complete enough experimental details.

**2.C. Evaluation of Experimental Temperature Dependence for Neutrals.** The temperature dependence of the free energies of solvation may be computed as

$$\Delta \Delta G_S^* = \Delta G_S^*(T) - \Delta G_S^*(T_0) \quad (16)$$

where  $\Delta G_S^*(T)$  is the experimental free energy of solvation at a given temperature  $T$ , and  $T_0 = 298$  K. To tabulate  $\Delta \Delta G_S^*$ , one requires a value of  $\Delta G_S^*(T_0)$  that is consistent with the  $\Delta G_S^*(T)$  data. In some cases, however, 298 K was not among the available temperatures in the dataset produced by the steps in sections 2.A and 2.B. Furthermore, even when available,  $\Delta G_S^*$  at any one temperature, such as  $T_0$ , is subject to experimental noise that can introduce a systematic error in the computed temperature dependence of the free energies of solvation that is significantly larger than the average experimental error in the temperature dependence of free energies of solvation. To make the data as consistent as possible, we always obtained  $\Delta G_S^*(T_0)$  by fitting all of the retained experimental data for a given compound,  $i$ , to

$$\Delta G_S^*(T, i) = a_i + b_i(T - T_0) + c_i[(T - T_0) - T \ln(T/T_0)] \quad (17)$$

where  $a_i$ ,  $b_i$ , and  $c_i$  are fitting parameters. The form of eq 17 is motivated by eq 1, which yields

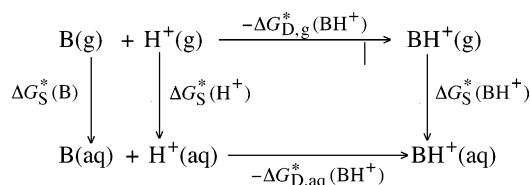
$$a_i = \Delta G_S^*(T_0) \quad (18)$$

The final result of this analysis is an extensive database of experimental data. A listing of the individual compounds, the number of experimental points used for each compound, the temperature range covered, and the sources for the data used in the database is available in the Supporting Information.

The current database is composed of 4501 points for 348 compounds; 382 of those points are at 298 K and the remaining 4119 points are in the ranges 273–297 K or 299–373 K. This database encompasses a broad range of functionality, including alkanes, alkenes, alkynes, cycloalkanes, aromatics, alcohols, ethers, esters, carboxylic acids, nitrates, heterocyclic N-containing compounds, amines, nitriles, amides, fluorinated compounds, chlorinated compounds, brominated compounds, thiols, sulfides, and heterocyclic S-containing compounds. For a summary of all the data for solutes containing N, F, Cl, Br, or S in the dataset, see Table 1. For a description of the compounds in the dataset that contain only H, C, or O, see Chamberlin et al.<sup>26</sup>

**2.D. Evaluation of Experimental Temperature Dependence for Ions.** The computation of the free energy of solvation for ions requires the use of a thermodynamic cycle, as shown in Scheme 1 for anions and in Scheme 2 for cations. Using the cycles shown in Schemes 1 and 2, one can com-

## SCHEME 2



pute the standard-state free energies of solvation for anions,  $\Delta G_{\text{S}}^*(\text{A}^-)$ , and cations,  $\Delta G_{\text{S}}^*(\text{BH}^+)$ , respectively, as

$$\Delta G_{\text{S}}^*(\text{A}^-) = \Delta G_{\text{S}}^*(\text{AH}) + \Delta G_{\text{D,aq}}^*(\text{AH}) - \Delta G_{\text{S}}^*(\text{H}^+) - \Delta G_{\text{D,g}}^*(\text{AH}) \quad (19)$$

$$\Delta G_{\text{S}}^*(\text{BH}^+) = \Delta G_{\text{S}}^*(\text{B}) + \Delta G_{\text{S}}^*(\text{H}^+) - \Delta G_{\text{D,aq}}^*(\text{BH}^+) + \Delta G_{\text{D,g}}^*(\text{BH}^+) \quad (20)$$

where  $\Delta G_{\text{S}}^*(\text{AH})$  and  $\Delta G_{\text{S}}^*(\text{B})$ , are the standard-state free energies of solvation for the neutrals,  $\Delta G_{\text{D,g}}^*(\text{AH})$  and  $\Delta G_{\text{D,g}}^*(\text{BH}^+)$  are the standard-state gas-phase free energies of dissociation,  $\Delta G_{\text{D,aq}}^*(\text{AH})$  and  $\Delta G_{\text{D,aq}}^*(\text{BH}^+)$  are the standard-state aqueous free energies of dissociation, and  $\Delta G_{\text{S}}^*(\text{H}^+)$  is the standard-state free energy of solvation for the proton. To obtain the free energy of solvation as a function of temperature, however, it is also necessary to obtain enthalpies and heat capacities of solvation for the ions, which can be computed in an fashion analogous to the ionic free energies of solvation, yielding the following equations for the enthalpies of solvation for anions,  $\Delta H_{\text{S}}^*(\text{A}^-)$ , and cations,  $\Delta H_{\text{S}}^*(\text{BH}^+)$ .

$$\Delta H_{\text{S}}^*(\text{A}^-) = \Delta H_{\text{S}}^*(\text{AH}) + \Delta H_{\text{D,aq}}^*(\text{AH}) - \Delta H_{\text{S}}^*(\text{H}^+) - \Delta H_{\text{D,g}}^*(\text{AH}) \quad (21)$$

$$\Delta H_{\text{S}}^*(\text{BH}^+) = \Delta H_{\text{S}}^*(\text{B}) + \Delta H_{\text{S}}^*(\text{H}^+) - \Delta H_{\text{D,aq}}^*(\text{BH}^+) + \Delta H_{\text{D,g}}^*(\text{BH}^+) \quad (22)$$

The heat capacities of solvation for anions,  $\Delta C_{\text{p,S}}^*(\text{A}^-)$ , and cations,  $\Delta C_{\text{p,S}}^*(\text{BH}^+)$ , may be computed as

$$\Delta C_{\text{p,S}}^*(\text{A}^-) = \Delta C_{\text{p,S}}^*(\text{AH}) + \Delta C_{\text{p,D,aq}}^*(\text{AH}) - \Delta C_{\text{p,S}}^*(\text{H}^+) - \Delta C_{\text{p,D,g}}^*(\text{AH}) \quad (23)$$

$$\Delta C_{\text{p,S}}^*(\text{BH}^+) = \Delta C_{\text{p,S}}^*(\text{B}) + \Delta C_{\text{p,S}}^*(\text{H}^+) - \Delta C_{\text{p,D,aq}}^*(\text{BH}^+) + \Delta C_{\text{p,D,g}}^*(\text{BH}^+) \quad (24)$$

The thermodynamic properties used to compute the free energies, enthalpies, and heat capacities of solvation for propanoate anion and *n*-propyl ammonium cation are shown in Table 2, along with the appropriate references. When the thermodynamic properties of solvation for the ion are used in eq 14, ionic free energies of solvation as a function of temperature are derived. The resultant free energies of solvation for the *n*-propyl ammonium cation and propanoate anion along with those for the neutrals, *n*-propylamine and propanoic acid, are shown in Table 3.

## 3. Developing the Model

The SM8T model predicts the temperature-dependence of aqueous free energies of solvation for compounds containing N, F, S, Cl, or Br and is an extension of the SM8 aqueous

TABLE 1: Number of Free Energies of Solvation in Database<sup>a</sup>

	no. of data	no. of compounds
nitrates	61	9
heterocycles	220	25
amines	241	20
nitriles	117	11
amides	79	6
chlorine	842	51
chlorine-bromine	15	2
bromine	92	9
chlorine-fluorine	58	6
fluorine	146	12
thiols	38	4
sulfides and disulfides	124	9
dimethylsulfoxide	11	1
thiophenes	8	2
total	2052	167

<sup>a</sup> In this table, a solvation free energy for a given molecule at one temperature is counted as one datum.

solvation model reported elsewhere.<sup>28</sup> To adequately describe SM8T, it is necessary to first describe SM8 and our prior work on the development of SM6T. Finally, we will discuss SM8T.

**3.A. SM8.** SM8 is the most recent version of the SM<sub>x</sub> series of models developed by Cramer, Truhlar, and co-workers.<sup>3,56–59</sup> In SM8, the free energy of solvation is computed as the sum of two contributions

$$\Delta G_{\text{S}}^* = \Delta G_{\text{ENP}} + G_{\text{CDS}} \quad (25)$$

where  $\Delta G_{\text{ENP}}$  is the free energy contribution due to electrostatic interactions between the solute and solvent modeled as bulk solvent, and  $G_{\text{CDS}}$  is a parametrized term designed to account for first-solvation-shell effects and for approximations used in the calculation of  $\Delta G_{\text{ENP}}$ . In implicit solvation models, such as the SM<sub>x</sub> series, the bulk electrostatic contribution to the free energy of solvation is computed by approximating the solvent as a homogeneous continuous dielectric medium characterized by the experimental dielectric constant,  $\epsilon$ , of the solvent in question.

The bulk electrostatic contribution to the free energy of solvation may be computed in a number of ways.<sup>5–8</sup> In SM8,  $\Delta G_{\text{ENP}}$  is computed as<sup>28</sup>

$$\Delta G_{\text{ENP}} = \Delta E_{\text{EN}} + G_{\text{p}} \quad (26)$$

where  $\Delta E_{\text{EN}}$  is the change in the solute's internal free energy upon insertion into solution from the gas phase, and  $G_{\text{p}}$  is the free energy of polarization. In practice, we usually retain only the electronic contribution to  $\Delta E_{\text{EN}}$ , assuming that the vibrational–librational–rotational contributions are small enough to be neglected and also neglecting the change in solute geometry.

Our treatment of the solute in solution approximates the solute as a collection of atom-centered partial atomic charges located in an otherwise empty cavity composed of overlapping spheres centered on the nuclei of the solute. There has been much debate about the appropriate means to define the size of the spheres.<sup>60–62</sup> SM8 involves a set of intrinsic atomic radii, referred to as Coulomb radii, that are assumed to depend only on atomic number and the hydrogen bond acidity of the solvent; in particular, they are independent of atomic charge and hybridization state of the atoms of the solute. The usefulness of such radii has been validated by the success of the SM<sub>x</sub> models.



**TABLE 2: The Standard-State Free Energies (kcal/mol), Enthalpies (kcal/mol), and Heat Capacities (cal<sup>-1</sup> mol<sup>-1</sup> K<sup>-1</sup>) of Dissociation and Solvation in the Gas-Phase and Solution at 298 K**

Propanoic Acid						Reference <sup>a</sup>
$\Delta\Delta G_{\text{S}}^*(\text{AH})$	-6.46	$H_{\text{S}}^*(\text{AH})$	-13.49	$C_{\text{P,S}}^*(\text{AH})$	38.25	31
$\Delta G_{\text{D,aq}}^*(\text{AH})$	6.79	$H_{\text{D,aq}}^*(\text{AH})$	-0.17	$C_{\text{P,D,aq}}^*(\text{AH})$	-32.83	101
$\Delta G_{\text{D,g}}^*(\text{AH})$	340.05	$H_{\text{D,g}}^*(\text{AH})$	347.22	$C_{\text{P,D,g}}^*(\text{AH})$	3.67 <sup>b</sup>	102
$\Delta G_{\text{S}}^*(\text{H}^+)^c$	-265.52	$H_{\text{S}}^*(\text{H}^+)^c$	-265.59	$C_{\text{P,S}}^*(\text{H}^+)^c$	-21.02	103
$\Delta G_{\text{S}}^*(\text{A}^-)$	-72.31	$H_{\text{S}}^*(\text{A}^-)$	-90.94	$C_{\text{P,S}}^*(\text{A}^-)$	19.37	
<i>n</i> -Propylamine						
$\Delta\Delta G_{\text{S}}^*(\text{B})$	-4.39	$H_{\text{S}}^*(\text{B})$	-13.31	$C_{\text{P,S}}^*(\text{B})$	55.21	31
$\Delta G_{\text{D,aq}}^*(\text{BH}^+)$	14.42	$H_{\text{D,aq}}^*(\text{BH}^+)$	13.84	$C_{\text{P,D,aq}}^*(\text{BH}^+)$	7.60	42
$\Delta G_{\text{D,g}}^*(\text{BH}^+)$	211.30	$H_{\text{D,g}}^*(\text{BH}^+)$	212.79	$C_{\text{P,D,g}}^*(\text{BH}^+)$	4.71 <sup>b</sup>	104
$\Delta G_{\text{S}}^*(\text{H}^+)^c$	-265.52	$H_{\text{S}}^*(\text{H}^+)^c$	-265.59	$C_{\text{P,S}}^*(\text{H}^+)^c$	-21.02	103
$\Delta G_{\text{S}}^*(\text{BH}^+)^c$	-71.14	$H_{\text{S}}^*(\text{BH}^+)$	-75.62	$C_{\text{P,S}}^*(\text{BH}^+)$	28.03	

<sup>a</sup> Except where indicated otherwise, all three values were taken from the reference given in this column. <sup>b</sup> This property was computed using mPW1PW96 with the 6-31+G(d,p)<sup>97,98</sup> basis set using a frequency scaling factor of 0.9515.<sup>105</sup> <sup>c</sup> These properties were obtained using a fit to data for the free energy of solvation for H<sup>+</sup> at 298, 323, 348, and 373 K from Bartels et al.,<sup>103</sup> which is sufficient for the calculation of free energies of solvation at those points but should be treated with caution when interpolating or extrapolating to other temperatures.

**TABLE 3: Free Energy of Solvation (kcal/mol) for the Proton, H<sup>+</sup>; Propanoate Anion, A<sup>-</sup>; Propanoic Acid, AH; *n*-Propyl Ammonium Cation, BH<sup>+</sup>; and *n*-Propylamine, B, as a Function of Temperature (K)**

temp	$\Delta G_s^*(\text{H}^+)^a$	$\Delta G_s^*(\text{A}^-)$	$\Delta G_s^*(\text{AH})$	$\Delta G_s^*(\text{BH}^+)$	$\Delta G_s^*(\text{B})$
298.15	-265.52	-72.31	-6.46	-71.14	-4.39
323.15	-265.51	-70.76	-5.91	-70.79	-3.69
348.15	-265.45	-69.26	-5.44	-70.50	-3.11
373.15	-265.36	-67.79	-5.03	-70.26	-2.62

<sup>a</sup> These values were taken from Bartels et al.<sup>103</sup>

The free energy of polarization is the free energy change associated with the favorable solute–solvent electrostatic interactions that form upon insertion of the solute into the solvent minus the cost of the accompanying distortion of the solvent molecules and solvent structure from the pure solvent state. The electrostatic interactions between the solute and the surrounding dielectric medium are calculated using the generalized Born equation<sup>63–66</sup>

$$G_p = -\frac{1}{2} \left( 1 - \frac{1}{\epsilon} \right) \sum_{k,k'} q_k q_{k'} \gamma_{kk'} \quad (27)$$

where the right-hand side of the equation is a double sum over all the atoms  $k$  and  $k'$  in the solute,  $q_k$  is the partial atomic charge of atom  $k$  computed using charge model 4<sup>27,67,68</sup> (CM4),  $\epsilon$  is the experimental dielectric constant of the solvent, and  $\gamma_{kk'}$  is the Coulomb integral between atoms  $k$  and  $k'$ . The Coulomb integral is calculated as<sup>65</sup>

$$\gamma_{kk'} = \frac{1}{\sqrt{r_{kk'}^2 + \alpha_k \alpha_{k'} \exp(r_{kk'}^2/d_{kk'}^2 \alpha_k \alpha_{k'})}} \quad (28)$$

where  $r_{kk'}$  is the distance between atoms  $k$  and  $k'$ ,  $d_{kk'}$  is a parameter, and  $\alpha_k$  is an effective Born radius for atom  $k$  that depends on the solute geometry and the intrinsic Coulomb radius of each atom in the solute. The bulk electrostatic contribution to the free energy of solvation computed using the generalized Born equation is a free energy.<sup>69</sup>

The change,  $\Delta E_{\text{EN}}$ , in internal energy upon insertion of the solute into liquid solution from the gas phase has two contributions:  $\Delta E_{\text{E}}$ , which is the energy required to distort the solute electronic structure from its gas-phase optimum, and  $\Delta E_{\text{N}}$ , which is the cost associated with distortion of the solute geometry from its gas-phase optimum. The solute internal electronic,  $\Delta E_{\text{E}}$ , and solute–solvent polarization,  $G_p$ , components of  $\Delta G_{\text{ENP}}$  are

**TABLE 4: Parameters in SM6T (ccal Å<sup>-2</sup> mol<sup>-1</sup> K<sup>-1</sup>)**

	6	8	161	181	661	662	861
B Coefficient							
final SM6T model	2.18	2.18	29.47	71.95	2.18	2.18	2.18
C Coefficient							
final SM6T model	-84.45	-84.45	78.59	166.23	78.59	-266.91	166.23

minimized in a self-consistent reaction field<sup>6,70</sup> calculation through inclusion of an additional term in the Fock operator.

The energy change associated with distortion of the solute geometry,  $\Delta E_{\text{N}}$ , can be calculated by optimizing the solute geometry in solution. This latter process can be relatively time-consuming, since the solute electronic structure is iteratively computed for each step in the optimization of the solute geometry in solution. One option that has been explored in previous work<sup>71</sup> is to use the solute gas-phase geometry and to allow only the electronic structure to relax in solution. We have found that this approximation yields reliable results for most cases.<sup>71</sup>

As mentioned above,  $G_{\text{CDS}}$  is an empirically derived term designed to minimize the deviation between experiment and  $\Delta G_{\text{ENP}}$ . It is intended to account for solvation phenomena that are not included in the bulk electrostatic term. It includes the free energy cost associated with forming a solute-shaped cavity in the solvent; changes in solvent structure beyond those included in  $\Delta E_{\text{EN}}$ ; non-bulk electrostatic phenomena, such as dispersion and Pauli exchange-repulsion; and all of the approximations (such as the choice of Coulomb radii) used in the calculation of  $\Delta G_{\text{ENP}}$ . Previous studies<sup>72–75</sup> have shown that the free energies associated with these phenomena are to a good approximation functions of the solvent accessible surface areas<sup>76–79</sup> (SASAs) of the solute atoms. Note that the atomic radii used to define the SASA for the  $G_{\text{CDS}}$  term are *not* the same as the Coulomb radii. In particular, we use the radii of Bondi<sup>80</sup> for SASA calculations. The SASA calculation also

**TABLE 5: Comparison of the Variation Relative to the Value at 298 K of Free Energies of Solvation,  $\Delta\Delta G_S^*$  (kcal/mol), and Bulk Electrostatic,  $\Delta\Delta G_{\text{ENP}}$  (kcal/mol) and Nonbulk Electrostatic,  $\Delta G_{\text{CDS}}$  (kcal/mol), Components Computed Using SM8T to Experiment at Various Temperatures, K, for Ionic and Neutral Compounds**

temp	experiment	SM8T			experiment	SM8T		
	$\Delta\Delta G_S^*$	$\Delta\Delta G_{\text{ENP}}$	$\Delta G_{\text{CDS}}$	$\Delta\Delta G_S^*$		$\Delta\Delta G_{\text{ENP}}$	$\Delta G_{\text{CDS}}$	$\Delta\Delta G_S^*$
Propanoate Anion					Propanoic Acid			
298.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
323.15	1.54	0.13	0.34	0.46	0.55	0.01	0.48	0.49
348.15	3.05	0.26	0.61	0.87	1.03	0.02	0.88	0.90
373.15	4.52	0.42	0.83	1.25	1.43	0.03	1.21	1.24
<i>n</i> -Propyl Ammonium Cation					<i>n</i> -Propylamine			
298.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
323.15	0.35	0.12	0.66	0.77	0.69	0.00	0.65	0.65
348.15	0.64	0.25	1.23	1.48	1.28	0.01	1.22	1.23
373.15	0.88	0.39	1.74	2.13	1.76	0.01	1.72	1.73

**TABLE 6: Mean Signed Error (MSE), Mean Unsigned Error (MUE), and Root Mean Squared Error (RMSE) for Different Parametrizations of SM8T for Compounds Containing N (kcal/mol)**

	MSE	MUE	RMSE	MSE	MUE	RMSE
Molecular Coefficients			Null Hypothesis			
nitrates	0.01	0.01	0.02	-0.19	0.32	0.37
heterocycles	0.00	0.01	0.01	-0.78	0.96	1.12
amines	0.00	0.01	0.01	-0.71	0.95	1.12
nitriles	0.00	0.01	0.02	-0.26	0.41	0.47
amides	0.00	0.02	0.02	-0.40	0.56	0.66
average	0.00	0.01	0.02	-0.47	0.64	0.75
error for $T > 348$ K	0.01	0.02	0.04	-1.55	1.57	1.63
error for $T < 298$ K	0.00	0.02	0.04	0.45	0.45	0.55
Nitrogen Null Hypothesis			Unrestricted Nitrogen Model			
nitrates	0.02	0.07	0.09	0.05	0.07	0.09
heterocycles	-0.21	0.27	0.32	-0.02	0.09	0.11
amines	-0.21	0.29	0.35	0.00	0.08	0.09
nitriles	0.00	0.05	0.06	0.01	0.04	0.05
amides	-0.19	0.30	0.36	0.04	0.12	0.14
average	-0.12	0.20	0.23	0.02	0.08	0.10
error for $T > 348$ K	-0.46	0.49	0.64	0.00	0.15	0.22
error for $T < 298$ K	0.14	0.14	0.20	0.02	0.05	0.07
No-C Coefficient Nitrogen Model			Final Nitrogen Model			
nitrates	0.04	0.07	0.10	0.03	0.07	0.09
heterocycles	-0.02	0.09	0.11	-0.01	0.09	0.11
amines	0.00	0.08	0.09	0.01	0.08	0.10
nitriles	0.02	0.04	0.06	0.02	0.04	0.06
amides	0.04	0.11	0.14	0.03	0.11	0.14
average	0.02	0.08	0.10	0.02	0.08	0.10
error for $T > 348$ K	0.02	0.15	0.22	0.02	0.15	0.22
error for $T < 298$ K	0.04	0.05	0.09	0.04	0.06	0.09

requires a solvent radius, and this is taken as  $0.4 \text{ \AA}^{27,81}$  in SM6, SM8, SM6T, and SM8T.

The  $G_{\text{CDS}}$  term has the following form,<sup>4,27,82</sup>

$$G_{\text{CDS}} = \sum_k \sigma_k A_k \quad (29)$$

where the first sum is over all atoms  $k$  in the solute,  $A_k$  is the solvent-accessible surface area of atom  $k$ , and  $\sigma_k$  is a parametrized functional that is written as<sup>82</sup>

$$\sigma_k = \tilde{\sigma}_k + \sum_{m \neq k} \sum_{\tau} \tilde{\sigma}_{Z_k Z_m \tau} F_{Z_k Z_m \tau}(\{r_{km}\}) \quad (30)$$

where the sum is over all the atoms  $m$  in the molecule except  $k$ ; the sum over  $\tau$  is over one or two types of geometric factors,  $F_{Z_k Z_m \tau}$ ;  $\tilde{\sigma}_k$  is a coefficient that depends on the atomic number  $Z_k$  of atom  $k$ ; and  $\tilde{\sigma}_{Z_k Z_m \tau}$  is a coefficient that depends on the

atomic numbers  $Z_k$  and  $Z_m$  of atoms  $k$  and  $m$  and the type of  $F_{Z_k Z_m \tau}$ . The functional  $F_{Z_k Z_m \tau}$  is designed to account for the dependence of atomic properties on the nature of the functional group in which the atom appears;  $F_{Z_k Z_m \tau}$  depends on the set of all the internuclear distances  $\{r_{km}\}$  in the molecule. For compounds containing H, C, N, O, F, S, Cl, or Br, SM8 has 8  $\sigma_Z$  coefficients— $\tilde{\sigma}_1, \tilde{\sigma}_6, \tilde{\sigma}_7, \tilde{\sigma}_8, \tilde{\sigma}_9, \tilde{\sigma}_{16}, \tilde{\sigma}_{17}$ , and  $\tilde{\sigma}_{35}$ —corresponding, respectively, to H, C, N, O, F, S, Cl, and Br, and it has 15  $\tilde{\sigma}_{ZZ\tau}$  coefficients  $\tilde{\sigma}_{161}, \tilde{\sigma}_{171}, \tilde{\sigma}_{172}, \tilde{\sigma}_{181}, \tilde{\sigma}_{1(16)1}, \tilde{\sigma}_{661}, \tilde{\sigma}_{671}, \tilde{\sigma}_{662}, \tilde{\sigma}_{761}, \tilde{\sigma}_{762}, \tilde{\sigma}_{763}, \tilde{\sigma}_{861}, \tilde{\sigma}_{871}, \tilde{\sigma}_{881}$ , and  $\tilde{\sigma}_{(16)(16)1}$ .

An additional parameter,  $\tilde{\sigma}_{8(16)1}$ , was also created during the course of this project. The parameter was designed to account for the unique effect of O—S bonds in sulfoxides upon the temperature dependence of the free energy of solvation. The development of this new parameter also entailed the development of a new functional form,  $F_{8(16)1}$ , for a geometric factor. As with the other geometric factors,<sup>61,83–89</sup>  $F_{8(16)1}$  is a geometry-dependent switching function called a cutoff tanh function. The general form for this function is

$$F(r_{km} | \bar{R}_{Z_k Z_m}, \Delta R) = \begin{cases} \exp\left[\frac{\Delta R}{\Delta R - r_{km} + \bar{R}_{Z_k Z_m}}\right] & r_{km} \leq \bar{R}_{Z_k Z_m} + \Delta R \\ 0 & \text{otherwise} \end{cases} \quad (31)$$

where  $\Delta R$  is the width or the range over which the function switches, and  $\bar{R}_{Z_k Z_m}$  is the midpoint. For the particular case of  $F_{8(16)1}$ , we took  $\Delta R = 0.30 \text{ \AA}$ , and  $\bar{R}_{Z_k Z_m} = 1.71 \text{ \AA}$ .

An important point to note is that results calculated with SM8 are independent of the choice of electronic structure theory and basis set. In SM8T, we also take the temperature dependence to be independent of both the level of electronic structure theory chosen and the basis set.

**3.B. SM6T.** Solvation model 6 with temperature dependence (SM6T)<sup>26</sup> is designed to predict the temperature dependence of aqueous free energies of solvation for compounds containing H, C, and O in the range 273–373 K. It is based upon Solvation model 6 (SM6),<sup>27</sup> which is an aqueous solvation model designed to accurately reproduce the solvation free energies for neutral and ionic compounds. SM6T partitions the temperature dependence of the free energy of solvation into two components: the temperature dependence of the bulk electrostatic contribution to the free energy of solvation, which is computed using the generalized Born equation, and the temperature dependence of first-solvation-shell effects, which is modeled using a parametrized solvent-exposed surface-area dependent term (which will be described further in the next section). The final model used

**TABLE 7: Coefficients Used for Each Parametrization of SM8T for Compounds Containing N (ccal Å<sup>-2</sup> mol<sup>-1</sup> K<sup>-1</sup>)**

	7	171	671	761	762	763	871
<i>B</i> Coefficient							
unrestricted nitrogen model	26.55	12.79	−20.24	13.00	130.84	−25.70	3.15
nitrogen model with no C coefficient	−6.12	28.73	−10.43	13.32	112.54	15.79	2.79
final nitrogen model		33.65		10.20	63.04		
<i>C</i> Coefficient							
unrestricted nitrogen model	393.53	−204.37	−160.71	2.87	344.50	−449.16	1.75
nitrogen model with no C coefficient							
final nitrogen model							

a total of seven parameters to predict the temperature dependence of the free energy of solvation for a wide range of compounds containing H, C, and O. We found that SM6T predicts the temperature dependence of aqueous free energies of solvation with a mean unsigned error of 0.08 kcal/mol over our entire database of compounds containing H, C, and O. Although the aqueous parameters of SM8 are slightly different from those of SM6, the parameters determined in the development of SM6T continue to be used without change in SM8T. (See Table 4)

**3.C. SM8T.** In SM8T, we account for the temperature dependence of the free energy of solvation,  $\Delta\Delta G_S^*$ , in both the bulk electrostatic contribution and in the nonbulk electrostatic contribution,

$$\Delta\Delta G_S^* = \Delta\Delta G_{\text{ENP}}(T) + \Delta G(T) \quad (32)$$

where  $\Delta\Delta G_{\text{ENP}}$  is the temperature dependence associated with  $\Delta G_{\text{ENP}}$ , and  $\Delta G_{\text{CDS}}$  is an empirically derived term designed to account for the temperature dependence of first-solvation-shell phenomena and other approximations associated with use of the generalized Born equation. The first term on the right-hand side of eq 42 is calculated directly as

$$\Delta\Delta G_{\text{ENP}} = \Delta G_{\text{ENP}}(T) - \Delta G_{\text{ENP}}(T_0) \quad (33)$$

As described in section 3.B, the bulk electrostatic contribution to the free energy of solvation depends on the partial atomic charges of the solute, the Coulomb radii of the solute atoms, the solute geometry, and the dielectric constant of the solvent, all of which might, in principle, depend on temperature. As was discussed in our prior paper,<sup>26</sup> we compute  $\Delta\Delta G_{\text{ENP}}$  using the variation of the solvent dielectric constant as a function of temperature,  $\epsilon(T)$ . The temperature dependence of the dielectric constant reflects the ability of the bulk solvent to reorient and polarize in response to the solute charge density. For water, the temperature dependence of the dielectric constant is well approximated by a cubic polynomial.<sup>90</sup>

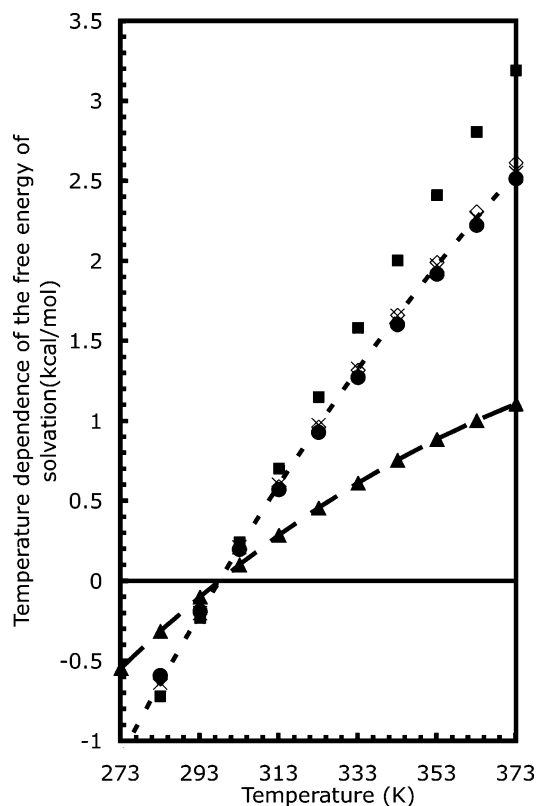
We found in our prior paper<sup>26</sup> and within our current dataset, as well, that the effect of the temperature dependence of the dielectric constant of water is very small. The largest effect we found was for urea in water, for which  $\Delta\Delta G_{\text{ENP}}$  varies by 0.09 kcal/mol over the temperature range of 273–373 K, over which range the dielectric constant of water varies from 87.90 to 55.58. The average variation of  $\Delta\Delta G_{\text{ENP}}$  over all of the compounds in the dataset is only 0.02 kcal/mol. We note, however, that this small effect becomes much larger in organic solvents, for example, in octanol, for which the dielectric constant varies from 11.92 to 5.32 over that same temperature range<sup>90</sup> and the corresponding values of  $\Delta\Delta G_{\text{ENP}}$  for urea in octanol vary by 1.25 kcal/mol. As a consequence, although we recognize that the effect of  $\Delta\Delta G_{\text{ENP}}$  is small in water, we account for it since it will have a more significant effect in organic solvents to which one might extend the model.

Additionally, the temperature dependence of the dielectric constant will play a markedly greater role in the solvation of ionic compounds. This particular issue is of interest because our initial work on ionic compounds indicates that ionic compounds do show a significant difference in the variation of the free energy of solvation relative their neutral counterparts (See Tables 3 and 5). However, the variation of the bulk electrostatic contribution will not account for the entire difference. Table 5 shows that the variation in the electrostatic contributions to the free energy of solvation vary by  $\sim 0.40$  kcal/mol for the two ions shown compared to  $\sim 0.02$  kcal/mol for their neutral counterparts, whereas the temperature dependence of the free energies of solvation for the ions differs from the neutrals by much more than that.

Another interesting point is that the temperature dependence of the free energy of solvation for the *n*-propyl ammonium cation is reduced relative to *n*-propylamine, whereas the temperature dependence of the propanoate anion is considerably increased relative to propanoic acid. SM8T, which has been developed for neutral compounds, does not address the difference between first-solvation-shell ion–solvent interactions and first-solvation-shell neutral–solvent interactions. Although SM8T successfully captures the qualitative effects of temperature upon the free energies of solvation, it does not quantitatively reproduce the experimental temperature dependence. SM8T does not, for instance, capture the reduction in the temperature dependence of the *n*-propyl ammonium cation relative to *n*-propylamine; neither does it predict an increased temperature dependence for the propanoate anion relative to propanoic acid. Accounting for these effects may entail a number of possibilities, such as including an explicit solvent molecule bound to the ionic solute, parametrizing the atomic radii of the solute as a function of temperature, or inclusion of additional surface tension coefficients that account for the unique effects of ions. The development of a temperature-dependent solvation model for ions, however, will be a very challenging task requiring careful analysis of the available literature on ions and development of a sufficiently robust database.

The bulk electrostatic contribution to the free energy of solvation,  $\Delta G_{\text{ENP}}$ , both at the reference temperature  $T_0$  and at the temperature of interest  $T$ , was computed using SM8 as it is implemented in MN-GSM<sup>91</sup> (a locally developed module of Gaussian), in GAMESSPLUS,<sup>92</sup> in SMXGAUSS<sup>93</sup> (a code that was developed within our group to allow use of our solvation models with Gaussian<sup>94</sup> style input), and in HONDOPLUS.<sup>95</sup> The  $\Delta G_{\text{ENP}}$  for each compound at each temperature was computed using the dielectric constant,  $\epsilon(T)$ , as represented by the polynomial equation reported in the *CRC Handbook of Chemistry and Physics*.<sup>90</sup> The electronic structure of each compound was optimized in the gas phase at the mPW1PW<sup>96</sup> level of theory using the 6-31+G(d,p)<sup>97,98</sup> basis set.

Since the temperature dependence of the bulk electrostatic contribution to the free energy of solvation in water is negligible, the  $\Delta G$  term incorporates almost all of the temperature



**Figure 1.** The temperature dependence,  $\Delta\Delta G_s^*$ , of the experimental (■) free energy of solvation for piperazine compared to various possible parametrizations of SM8T: null hypothesis (—) nitrogen null hypothesis (—▲—), nitrogen model with no-C coefficient (◇), unrestricted nitrogen model (—×—) and final nitrogen model (●).

dependence of the aqueous free energy of solvation in the SM8T model. This term may be written as

$$\Delta G_{\text{CDS}} = G_{\text{CDS}}(T) - G_{\text{CDS}}(T_0) \quad (34)$$

Comparing eq 34 to eq 1 suggests that we model this as

$$\Delta G_{\text{CDS}} = B(T - T_0) + C[(T - T_0) - T \ln(T/T_0)] \quad (35)$$

where  $B$  and  $C$  are parameters of the model. As in the SM8 model for  $\Delta G_{\text{CDS}}$ , we model  $\Delta G_{\text{CDS}}$  as proportional to solvent-exposed surface areas.<sup>6,61,88</sup> We applied the same functions and solvent-accessible surface areas used in eqs 29 and 30 to calculate  $B$  and  $C$

$$B = \sum_k \sigma_k^B A_k \quad (36)$$

$$C = \sum_k \sigma_k^C A_k \quad (37)$$

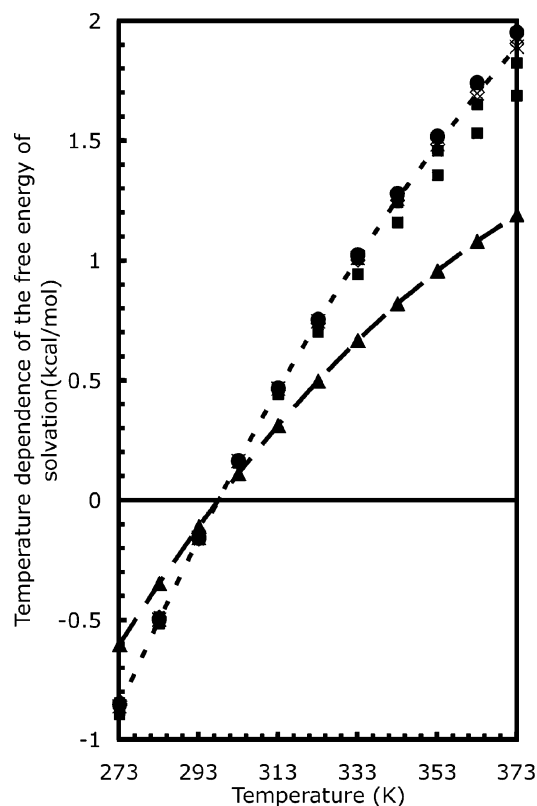
where  $\sigma_k^B$  and  $\sigma_k^C$  are parametrized as

$$\sigma_k = \tilde{\sigma}_{Z_k}^B + \sum_{m \neq k} \sum_{\tau} \tilde{\sigma}_{Z_k Z_m}^B F_{Z_k Z_m \tau}(\{r_{km}\}) \quad (38)$$

$$\sigma_k = \tilde{\sigma}_{Z_k}^C + \sum_{m \neq k} \sum_{\tau} \tilde{\sigma}_{Z_k Z_m}^C F_{Z_k Z_m \tau}(\{r_{km}\}) \quad (39)$$

where  $\tilde{\sigma}_{Z_k}^B$ ,  $\tilde{\sigma}_{Z_k}^C$ ,  $\tilde{\sigma}_{Z_k Z_m}^B$  and  $\tilde{\sigma}_{Z_k Z_m}^C$  are coefficients that depend on the atomic numbers,  $Z_k$  and  $Z_m$ , of atoms  $k$  and  $m$  respectively.

We did not include the  $F_{172}$  term in SM8T because it was designed<sup>99</sup> to account for the unique contributions of hydrazine-type functionality to the free energies of solvation. Hydrazines



**Figure 2.** The temperature dependence,  $\Delta\Delta G_s^*$ , of the experimental (■) free energy of solvation for butylamine as compared to different parametrizations of SM8T (see Figure 1 for labels).

are often explosive, and perhaps for that reason, few reports of the properties of these compounds at higher temperatures have appeared. We did not find any sources that would provide a sufficient number of temperatures to properly fit any hydrazine-containing compounds. Thus, there are potentially as many as 14 parameters for the N-containing compounds in SM8T: 7 for  $B$  and 7 for  $C$ , potentially 6 parameters for halogen-containing compounds, and potentially 8 parameters for the S-containing compounds. However, the introduction of 28 extra parameters to fit the temperature dependence of the free energy of solvation is excessive. We found through careful analysis a set of 8 parameters that, when combined with the 7 parameters used to fit the compounds that contain exclusively H, C, and O, quite accurately fits the temperature dependence of the free energy of solvation over the dataset. The analysis and selection of those parameters is discussed in detail in section 3.E.

**3.D. Parameter Optimization.** Our optimization of the  $\Delta G$  parameters during the development, testing, and finalization of the SM8T model minimized an error function. The parameters for the N-containing compounds, the halogen-containing compounds, and the S-containing compounds were each optimized separately. Two different types of error functions were used in the extension of SM8T to N, F, Cl, Br, and S-containing compounds. In the case of the N-containing compounds, a weighted error function based on classes of N-containing compounds was used. The error function was defined as

$$D_N = \sum_{i=1}^5 \sum_{m=1}^{s_i} \times \left[ \frac{1}{5n_m s_i} \sum_i^{n_m} (\Delta G_{\text{CDS}}(\text{Exp}; i) - \Delta G_{\text{CDS}}(\text{Calc}; i))^2 \right]^{1/2} \quad (40)$$



**TABLE 8: Mean Signed Error (MSE), Mean Unsigned Error (MUE), and Root Mean Squared Error (RMSE) for Different Parametrizations of SM8T for Compounds Containing F, Cl, and/or Br (kcal/mol)**

	MSE	MUE	RMSE	MSE	MUE	RMSE
Molecular Coefficients			Null Hypothesis			
chlorine	0.00	0.03	0.03	−0.04	0.34	0.42
chlorine−bromine	0.00	0.02	0.02	0.04	0.20	0.24
bromine	0.00	0.01	0.01	0.06	0.29	0.36
chlorine−fluorine	0.00	0.02	0.03	−0.23	0.42	0.54
fluorine	0.00	0.01	0.01	−0.35	0.56	0.66
average	0.00	0.02	0.02	−0.10	0.36	0.45
error for $T > 348$ K	0.00	0.02	0.03	−1.09	1.09	1.15
error for $T < 298$ K	0.00	0.03	0.04	0.29	0.29	0.36
Halogen Null Hypothesis			Unrestricted Halogen Model			
chlorine	−0.03	0.17	0.22	0.00	0.09	0.11
chlorine−bromine	0.04	0.18	0.22	0.02	0.03	0.05
bromine	0.01	0.13	0.16	0.00	0.03	0.04
chlorine−fluorine	−0.20	0.38	0.48	−0.01	0.10	0.14
fluorine	−0.22	0.35	0.40	0.02	0.07	0.10
average	−0.08	0.24	0.30	0.01	0.07	0.09
error for $T > 348$ K	−0.65	0.67	0.79	0.02	0.21	0.29
error for $T < 298$ K	0.17	0.18	0.24	0.00	0.06	0.09
No-C Coefficient Halogen Model			Final Halogen Model			
chlorine	0.01	0.09	0.11	0.00	0.09	0.11
chlorine−bromine	0.03	0.05	0.06	0.02	0.04	0.05
bromine	0.02	0.04	0.05	0.01	0.03	0.04
chlorine−fluorine	0.00	0.10	0.15	−0.02	0.10	0.14
fluorine	0.02	0.09	0.12	−0.01	0.08	0.11
average	0.02	0.07	0.10	0.00	0.07	0.09
error for $T > 348$ K	0.06	0.22	0.29	−0.01	0.21	0.29
error for $T < 298$ K	0.03	0.07	0.11	0.00	0.06	0.09

where the first sum is over all 5 classes,  $t$ , of N-containing compounds in the database; the second sum is over all  $s_t$  compounds,  $m$ , in a given class,  $t$ , and the third sum is over all  $n_m$  experimental data,  $i$ , for a given compound  $m$ .  $\Delta G_{\text{CDS}}(\text{Calc};i)$  is calculated using eqs 36–39, and  $\Delta G_{\text{CDS}}(\text{Exp};i)$  is the experimental value obtained by

$$\Delta G_{\text{CDS}}(\text{Exp}) = \Delta \Delta G_{\text{S}}^*(T) - \Delta \Delta G_{\text{ENP}}(T) \quad (41)$$

where the first term on the right-hand side is obtained from eqs 16 and 17, and the second term is calculated by the generalized Born approximation with the temperature-dependent dielectric constant of water. This error function was used in an effort to prevent certain key classes of compounds, namely, the nitrates and amines, from disproportionately affecting the model relative to other less represented classes of compounds, such as the amides and the nitriles.

The error function in eq 40 weights each class of compounds equally with the other classes, in which each compound in a class is weighted equally with the other compounds in the class. Unfortunately the danger of this approach is that classes with a very small number of compounds may skew the parametrization. For this reason, a different error function was used in the optimization of the parameters for the compounds containing F, S, Cl, and Br, for which some distinct classes contain only one or two compounds. A more appropriate error function for the halogen dataset has the following form

$$D_{\text{F,Cl,Br}} = \sum_{m=1}^{80} \times \left[ \frac{1}{80n_m} \sum_i^{n_m} (\Delta G_{\text{CDS}}(\text{Exp};i) - \Delta G_{\text{CDS}}(\text{Calc};i))^2 \right]^{1/2} \quad (42)$$

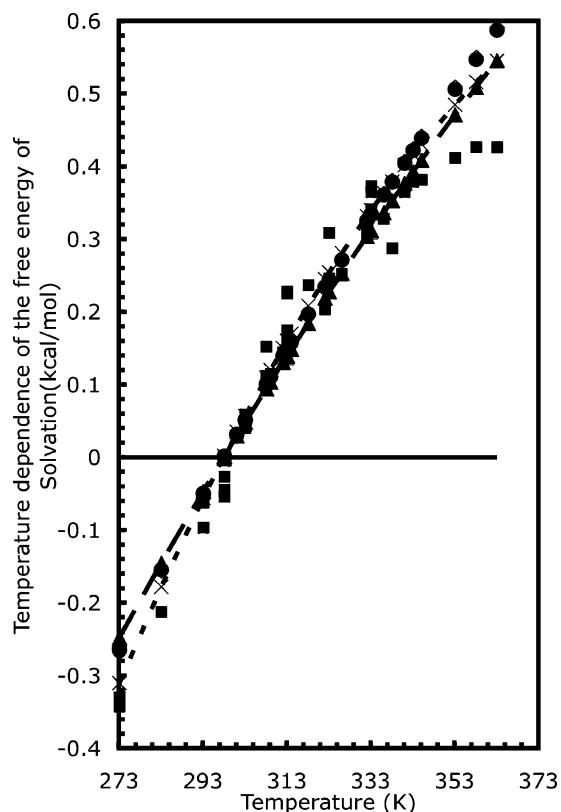
where the outer sum is over all 80 halogen-containing compounds in the database. This error function was used because some of the potential classes of compounds in the dataset, namely the chlorobromo compounds, had too few data to allow confident use of the error function in eq 40. Furthermore, there was considerably less interdependence of the parameters in the case of the halogen-containing compounds than the N-containing compounds, so it was less necessary to ensure that notable classes of compounds were properly represented in the dataset. The parameters for the N-containing compounds and the parameters for the halogen-containing compounds were optimized separately. Since only one compound, chloropicrin, contains both Cl and N, it was deemed unnecessary to reoptimize the parameters for N-containing compounds including this compound. Instead, only the halogen parameters were optimized including this compound. The parameters for the final N model were used to represent the temperature dependence introduced by the nitro group in chloropicrin.

The error function for S-containing compounds had the following form:

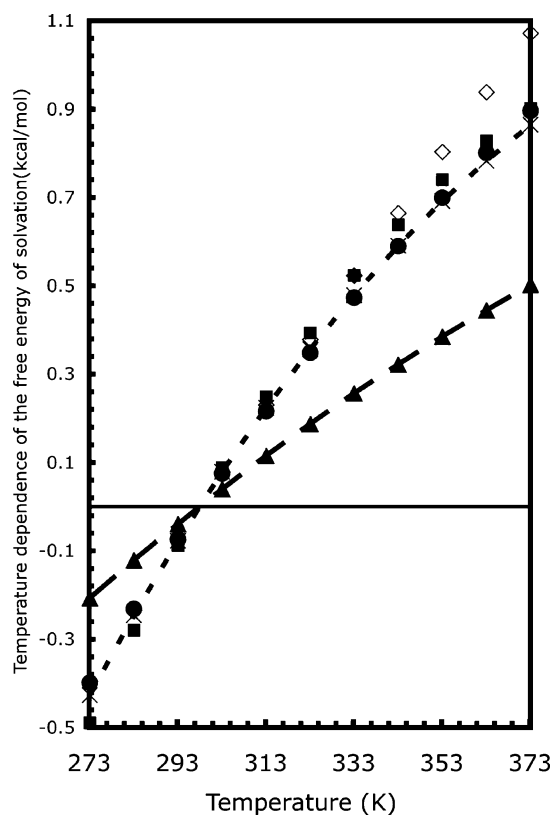
$$D_{\text{S}} = \sum_{m=1}^{16} \left[ \frac{1}{16n_m} \sum_i^{n_m} (\Delta G_{\text{CDS}}(\text{Exp};i) - \Delta G_{\text{CDS}}(\text{Calc};i))^2 \right]^{1/2} \quad (43)$$

where the sum is over all 16 compounds in the dataset. As with  $D_{\text{F,Cl,Br}}$ , the error function for the S-containing compounds was chosen to prevent less well-represented classes of compounds from skewing the parameters. For the S-containing compounds this challenge arose from the S-containing heterocycles. The entire class is represented by only two compounds, thiophene and 2-methylthiophene.

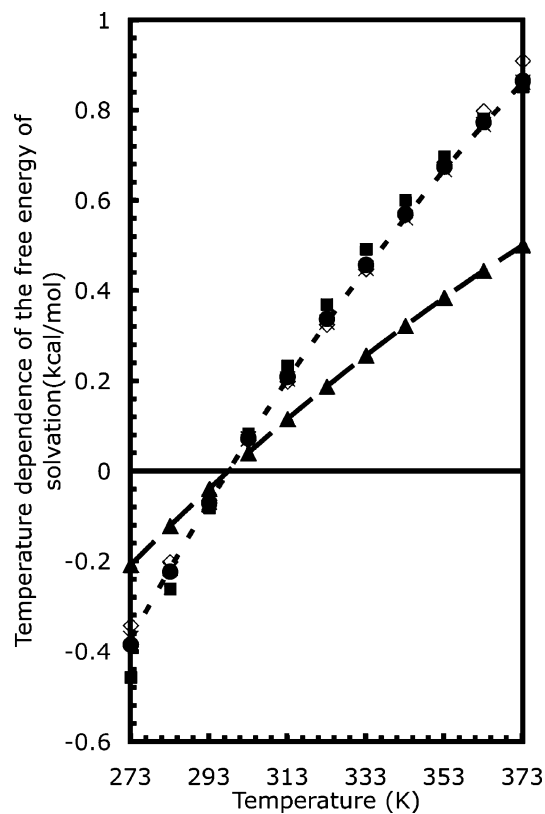
The error functions  $D_{\text{N}}$ ,  $D_{\text{F,Cl,Br}}$ , and  $D_{\text{S}}$  were minimized using the generalized reduced gradient code (GRG2) of Lasdon et al.<sup>100</sup> as implemented in Microsoft Excel 2004. In addition, the



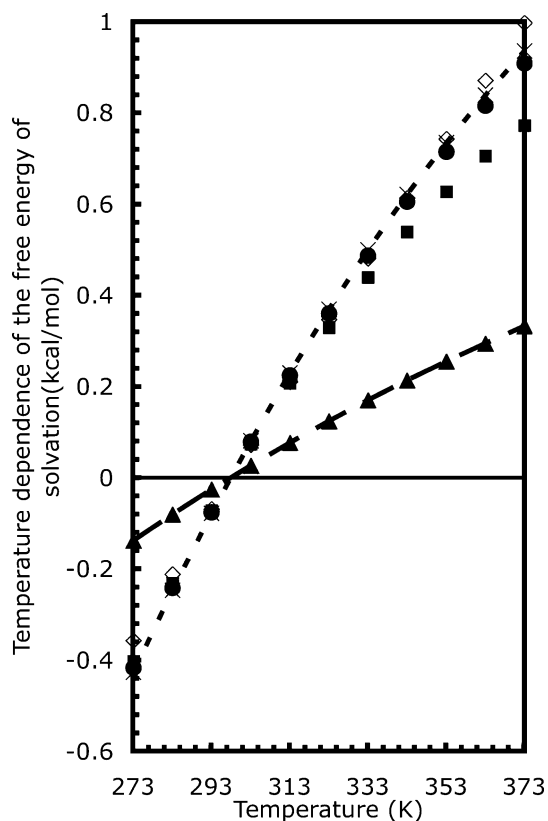
**Figure 3.** The temperature dependence,  $\Delta\Delta G_s^*$ , of the experimental (■) free energy of solvation for acrylonitrile as compared to different parametrizations of SM8T (see Figure 1 for labels).



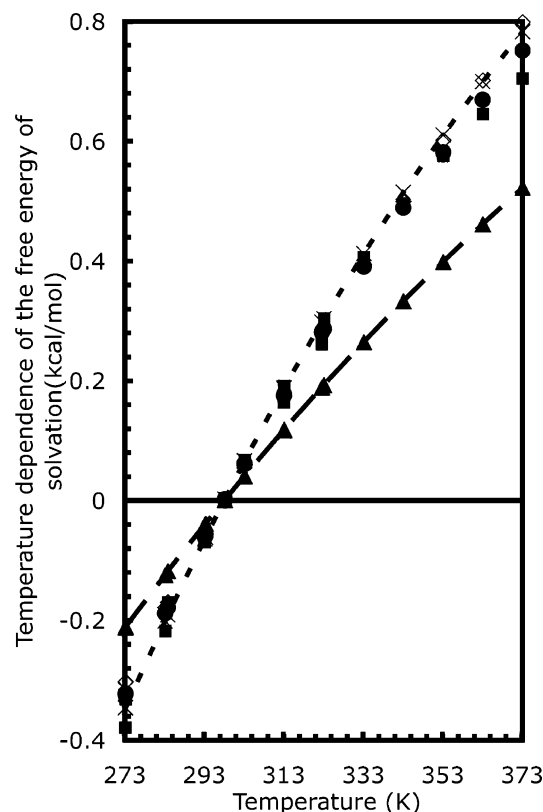
**Figure 5.** The temperature dependence,  $\Delta\Delta G_s^*$ , of the experimental (■) free energy of solvation for bromomethane as compared to different parametrizations of SM8T (see Figure 4 for labels).



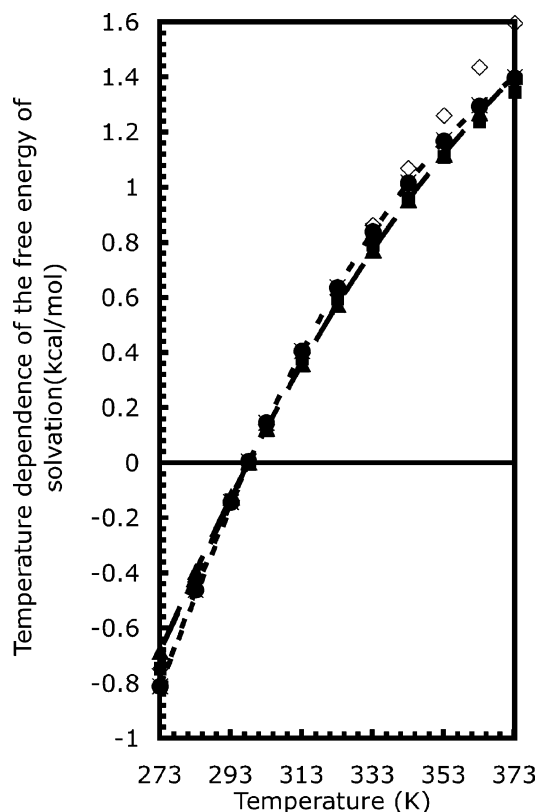
**Figure 4.** The temperature dependence,  $\Delta\Delta G_s^*$ , of the experimental (■) free energy of solvation for chloromethane as compared to different parametrizations of SM8T: null hypothesis (—), halogen null hypothesis (—▲—), unrestricted halogen model (—×—), halogen model with no C coefficient (◇), and final halogen model (●).



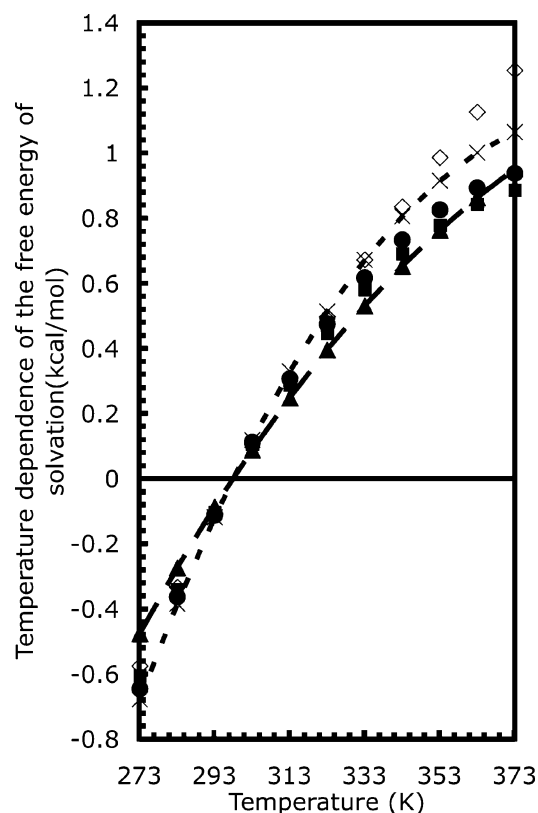
**Figure 6.** The temperature dependence,  $\Delta\Delta G_s^*$ , of the experimental (■) free energy of solvation for chlorofluoromethane as compared to different parametrizations of SM8T (see Figure 4 for labels).



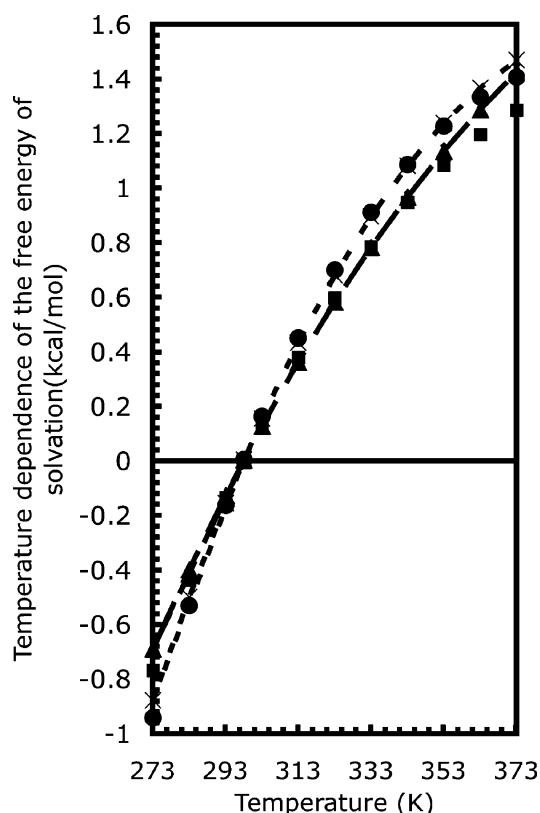
**Figure 7.** The temperature dependence,  $\Delta\Delta G_S^*$ , of the experimental (■) free energy of solvation for fluoromethane as compared to different parametrizations of SM8T (see Figure 4 for labels).



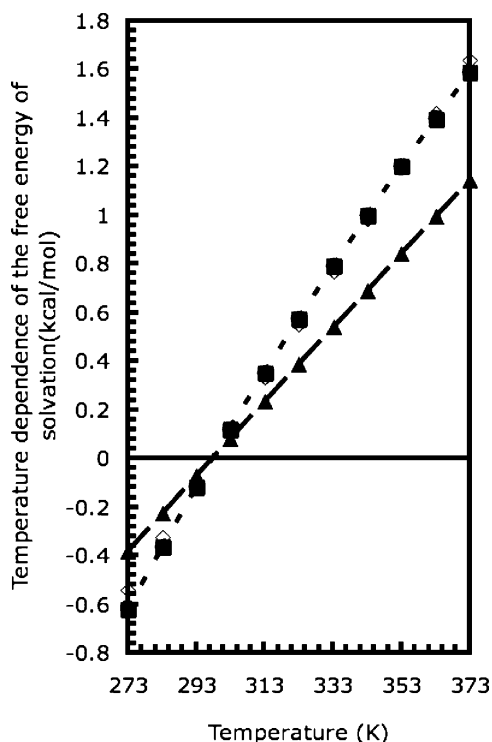
**Figure 9.** The temperature dependence,  $\Delta\Delta G_S^*$ , of the experimental (■) free energy of solvation for diethyl sulfide as compared to different parametrizations of SM8T (see Figure 8 for labels).



**Figure 8.** The temperature dependence,  $\Delta\Delta G_S^*$ , of the experimental (■) free energy of solvation for propanethiol as compared to different parametrizations of SM8T: null hypothesis (—), sulfur null hypothesis (—▲—), unrestricted sulfur model (—×—), sulfur model with no C coefficient (◇), and final sulfur model (●).



**Figure 10.** The temperature dependence,  $\Delta\Delta G_S^*$ , of the experimental (■) free energy of solvation for diethyl disulfide as compared to different parametrizations of SM8T (see Figure 8 for labels).



**Figure 11.** The temperature dependence  $\Delta\Delta G_s^*$  of the experimental (■) free energy of solvation for dimethylsulfoxide as compared to different parametrizations of SM8T (see Figure 8 for labels).

potential utility of the individual parameters for N-containing compounds was analyzed using a NAG linear optimization scheme.

Another measure of the effectiveness of the solvation model is to consider the accuracy of the model for a specific range of temperatures using the same error form for the error function as in eq 43. Table 6 shows the error of different models for data found within the ranges of 273–298 K and 348–373 K. The errors over the more restricted temperature ranges track equally well with the error functions shown in eqs 41–43, so they were not used due to concerns about whether parametrizations using them would be as sensitive to the curvature of the experimental data as parametrizations fit to the entire database. Additionally, a number of compounds in the dataset do not have representative points in one or both of these ranges, so there were concerns about the robustness of a model based upon the restricted temperature ranges. Nonetheless, they are helpful indicators of just how well the models fit the upper and lower limits of the temperature range.

**3.E. Parametrizations.** Having chosen a functional form for the predictive model and an appropriate error function, we proceeded to determine which parameters— $\sigma_{Z_k}^B$ ,  $\tilde{\sigma}_{Z_k}^C$ ,  $\tilde{\sigma}_{Z_k Z_m}^B$ , and  $\tilde{\sigma}_{Z_k Z_m}^C$ —are necessary to accurately reproduce the experimental temperature dependence of the aqueous free energies of solvation. Parameters for N-containing, halogen-containing, and S-containing compounds were developed independently, on the basis of the parameters previously developed in SM6T<sup>27</sup> for H, C, and O-containing compounds. The extension of SM8T was carried out holding the parameters for H, C, and O fixed at the values reported in our prior work.<sup>26</sup>

The parametrization of the SM8T model was divided into three steps: parametrization of SM8T for N-containing compounds, parametrization for compounds containing F, Cl, or Br, and parametrization for S-containing compounds. In

**TABLE 9: Coefficients Used for Each Parametrization of SM8T for Compounds Containing F, Cl, or Br (ccal Å<sup>-2</sup> mol<sup>-1</sup> K<sup>-1</sup>)**

	9	17	35
B Coefficient			
unrestricted halogen model	18.72	14.67	17.22
halogen model with no C coefficient	14.05	13.06	16.84
final halogen model	15.63	15.63	15.63
C Coefficient			
unrestricted halogen model	46.97	26.68	55.83
halogen model with no C coefficient			
final halogen model	34.15	34.15	34.15

each step, we considered a number of test parametrizations including:

1. where all the parameters are set equal to zero (“null hypothesis”);
2. where all the new parameters for N, F, S, Cl, and Br are set equal to zero, but the H, C, and O parameters are retained (“nitrogen/halogen/sulfur null hypothesis”);
3. where all of the available parameters relevant to the N, F, S, Cl, and Br atoms are used (“unrestricted nitrogen/halogen/sulfur”); and
4. where a unique pair of values for *B* and *C* is determined for each individual compound (“molecular coefficients”).

**3.E.1. Parametrization for the Nitrogen Dataset.** The use of molecular coefficients corresponds to the best-case scenario for any model that uses eq 17 to predict the temperature dependence of free energies of solvation. It can also be considered to give a rough estimate of the nonsystematic experimental error in the temperature dependence of the free energies of solvation in our database under the assumption that the effect of the temperature dependence of the heat capacity of solvation is small over the temperature range studied. Table 6 shows that such nonsystematic errors produce a mean unsigned error, MUE, of only 0.01 kcal/mol in the nitrogen dataset.

The null hypothesis, in contrast, corresponds to assuming that GCDS does not change as a function of temperature. Table 6 shows that this produces a mean unsigned error of 0.64 kcal/mol, which is considerably greater than experimental error in measuring free energies of solvation, which is estimated<sup>81</sup> as 0.2 kcal/mol. The error in the null hypothesis is, as it obviously should be, similar to the average variation in the aqueous free energies of solvation for all the compounds in the N dataset, which is 0.53 kcal/mol over a temperature range of 298–318 K. The “average variation” is simply the difference between the values of the free energy of solvation at 298 and 318 K for each compound in the dataset, averaged over the dataset. The range of 298–318 K was chosen because all of the compounds in the dataset either span that range or have endpoints within 5 K of 318 K, in which case an extrapolation using an independent fit to the experimental data gives a reliable estimate at 318 K.

The nitrogen null hypothesis assumes that the nitrogen atom has no effect on the temperature dependence of the free energy of solvation. This particular test was critical in determining whether extending SM8T to N-containing compounds would produce a worthwhile improvement. As Table 6 shows, assuming that nitrogen has no effect produces a mean unsigned error of 0.20 kcal/mol, which is quite significant.

Next, we determined the best fit to experiment that we could achieve using the available parameters. We performed a parametrization fit against the entire nitrogen database using all 14 possible coefficients:  $\tilde{\sigma}_7^B$ ,  $\tilde{\sigma}_{171}^B$ ,  $\tilde{\sigma}_{671}^B$ ,  $\tilde{\sigma}_{761}^B$ ,  $\tilde{\sigma}_{762}^B$ ,  $\tilde{\sigma}_{763}^B$ ,  $\tilde{\sigma}_{871}^B$ ,  $\tilde{\sigma}_7^C$ ,  $\tilde{\sigma}_{171}^C$ ,  $\tilde{\sigma}_{671}^C$ ,  $\tilde{\sigma}_{761}^C$ ,  $\tilde{\sigma}_{762}^C$ ,  $\tilde{\sigma}_{763}^C$ , and  $\tilde{\sigma}_{871}^C$ . The resulting fit



**TABLE 10: Mean Signed Error (MSE), Mean Unsigned Error (MUE), and Root Mean Squared Error (RMSE) for Different Parametrizations of SM8T for Compounds Containing S (kcal/mol)**

	MSE	MUE	RMSE	MSE	MUE	RMSE
Molecular Coefficients			Null Hypothesis			
thiols	0.00	0.05	0.06	−0.25	0.59	0.45
sulfides and disulfides	0.00	0.01	0.01	−0.22	0.45	0.55
sulfoxides	0.00	0.00	0.00	1.98	0.74	0.88
thiophenes	0.00	0.00	0.00	−0.42	0.42	0.57
average	0.00	0.02	0.02	0.27	0.55	0.61
error for $T > 348$ K	0.00	0.01	0.01	−1.02	1.42	1.07
error for $T < 298$ K	−0.01	0.04	0.06	0.26	0.28	0.34
Sulfur Null Hypothesis			No-C Coefficient Sulfur Model			
thiols	−0.03	0.11	0.13	0.03	0.11	0.13
sulfides and disulfides	0.05	0.10	0.13	0.06	0.08	0.11
sulfoxides	−0.16	0.23	0.27	0.01	0.03	0.03
thiophenes	−0.18	0.18	0.24	−0.12	0.12	0.17
average	−0.08	0.15	0.19	0.00	0.08	0.11
error for $T > 348$ K	−0.09	0.18	0.24	0.12	0.15	0.20
error for $T < 298$ K	0.10	0.12	0.16	0.04	0.08	0.12
Unrestricted Sulfur Model			Final Sulfur Model			
thiols	0.01	0.10	0.12	−0.01	0.09	0.11
sulfides and disulfides	0.03	0.06	0.08	0.02	0.07	0.09
sulfoxides	0.00	0.00	0.00	0.00	0.00	0.00
thiophenes	−0.13	0.13	0.20	−0.13	0.13	0.20
average	−0.02	0.07	0.10	−0.03	0.07	0.10
error for $T > 348$ K	0.05	0.11	0.14	−0.01	0.13	0.18
error for $T < 298$ K	−0.01	0.07	0.09	−0.01	0.07	0.10

(unrestricted N parametrization) shown in Table 6 produces a mean unsigned error of 0.08 kcal/mol. Table 7 shows the values of the coefficients. The improvement, while notable, comes at the cost of far too many parameters.

Our next step was to develop a parametrization that used all seven of the available parameters for the  $B$  coefficient and none of the parameters for the  $C$  coefficient for the purpose of testing whether the  $C$  coefficient was necessary. Table 6 shows that this no- $C$  coefficient fit is quite accurate, with a mean unsigned error of 0.08 kcal/mol. Additionally when one inspects charts of the predicted temperature dependence of the aqueous free energies of solvation against the temperature and compares the fit to experimental values for some representative compounds, such as butylamine (see Figure 2), one immediately sees that it manages to capture not only the slope but also the curvature of the experimental data. Our next step was to determine which of the parameters associated with the  $B$  coefficient were significant.

Our final N model uses a total of three additional parameters (beyond the seven parameters used to reproduce the temperature dependence for compounds containing H, C, and O)— $\tilde{\sigma}_{171}^B$ ,  $\tilde{\sigma}_{761}^B$ , and  $\tilde{\sigma}_{762}^B$ —to predict the temperature dependence of the free energy of solvation for N-containing compounds. The final nitrogen model reproduces the experimental data with a mean unsigned error of 0.08 kcal/mol. Comparison of the final nitrogen model to the unrestricted one and to the no- $C$  coefficient nitrogen model on several representative compounds (Figures 1–3) clearly indicates that both the qualitative and quantitative aspects of the temperature dependence of the free energy of solvation for N-containing compounds are well reproduced.

**3.E.2. Parametrization for the Halogen Dataset.** The error associated with the use of molecular coefficients on the halogen database once again emphasizes that the experimental error associated with the temperature dependence of free energies of solvation is considerably smaller than the absolute experimental error in free energies of solvation. In the case of the halogen

database the mean unsigned error (0.02 kcal/mol) is slightly larger than the value seen for the N or the H, C, and O datasets.

The mean unsigned error for the null hypothesis is shown in Table 7 (0.36). Perhaps a more important measure is average variation for halogen-containing compounds over the range 298–318 K, which is 0.46 kcal/mol. Using the null hypothesis, the temperature dependence of experimental points for halogens above 348 K has an error of 1.09 kcal/mol. (See Table 8)

The mean unsigned error due to assuming that F, Cl, and Br have no effect on the temperature dependence of the free energy of solvation, the halogen null hypothesis, is 0.24 kcal/mol. The sole use of the H, C, and O parameters manages to capture only one-third of the temperature dependence of the free energy of solvation, and inclusion of parameters for the halogens is critical to accurately reproduce the temperature dependence of free energies of solvation for halogenated compounds.

Our next step was to determine the maximum possible accuracy that could be achieved using only the available SMx parameters for F, Cl, and Br, which for these atoms consist of only the single-atom parameters,  $\tilde{\sigma}_9^B$ ,  $\tilde{\sigma}_{17}^B$ ,  $\tilde{\sigma}_{35}^B$ ,  $\tilde{\sigma}_9^C$ ,  $\tilde{\sigma}_{17}^C$ , and  $\tilde{\sigma}_{35}^C$ . The use of six parameters for the unrestricted halogen parametrization produces a mean unsigned error of only 0.07 kcal/mol. Although this was a pleasantly accurate result for only six parameters, we were dubious as to whether they were all necessary.

Our first consideration while developing this model was whether it was necessary to account for the unique effects of halogens upon the curvature of the temperature dependence of the free energy of solvation (no  $C$  coefficient model). As Figures 4–7 clearly show, the no  $C$  coefficient model is incapable of qualitatively reproducing the temperature dependence of the free energy of solvation.

Our next consideration was to combine parameters. We used a similar approach when working with the parametrization of SM8T for H, C, and O-containing compounds with good results. In this case the approach is justifiable on the basis of the observation that F, Cl, and Br are all halogens and are likely to interact with the solvent in a similar fashion. We combined all

**TABLE 11: Coefficients Used for Each Parametrization of SM8T for S-Containing Compounds (ccal Å<sup>-2</sup> mol<sup>-1</sup> K<sup>-1</sup>)**

	16	1161	8161	16161
B Coefficient				
unrestricted model	14.67	12.74	42.32	−5.01
final model	14.59		42.03	
C Coefficient				
unrestricted model	127.38	−17.23	9.67	−60.33
final model	129.62			

of the parameters for the *B* coefficient ( $\sigma_9^B$ ,  $\sigma_{17}^B$ , and  $\sigma_{35}^B$ ) into a single parameter,  $\tilde{\sigma}_{b1}^B$ , and all of the parameters for the *C* coefficient ( $\sigma_9^C$ ,  $\sigma_{17}^C$ , and  $\sigma_{35}^C$ ) into a single parameter,  $\tilde{\sigma}_{c1}^C$ . This produced a model, referred to as the final model, that agrees with experiment with a mean unsigned error of 0.07 kcal/mol (Table 9).

**3.E.3. Parametrization for the Sulfur Dataset.** The error associated with the use of molecular coefficients on the sulfur database as shown in Table 10, 0.03, is larger than the average errors for other types of compounds due to the much smaller number of experimental sources and number of compounds available. The null hypothesis when applied to the database of S-containing compounds produces an error of 0.55 kcal/mol. For the sulfur dataset, the average variation of the free energy of solvation from 298 to 318 K is 0.40 kcal. The observation that the temperature dependence of the free energy of solvation varies by 0.40 kcal/mol over 20 K, which is larger than the accuracy of the SM8 model, reemphasizes the importance of accounting for the temperature dependence of the free energy of solvation. The mean unsigned error associated with the sulfur null hypothesis is 0.15 kcal/mol.

The mean unsigned error associated with using the unrestricted sulfur model is 0.07 kcal/mol. Prior to the inclusion of the S–O parameter,  $\tilde{\sigma}_{8(16)1}$ , the unrestricted sulfur model, that is, using only the  $\tilde{\sigma}_{16}$ ,  $\tilde{\sigma}_{1(16)1}$ , and  $\tilde{\sigma}_{(16)(16)1}$  parameters, had a mean unsigned error of 0.09 kcal/mol, with the increased error attributable to dimethylsulfoxide data that could not be accurately reproduced without the inclusion of a parameter for sulfoxides.

During the development of the sulfur model, we considered a number of different approaches, including combining terms. We found that S–S bonds in disulfides did not uniquely affect the temperature dependence of the free energy of solvation, and as a consequence we discarded those parameters for the *B* and *C* coefficients,  $\tilde{\sigma}_{(16)(16)1}^B$  and  $\tilde{\sigma}_{(16)(16)1}^C$ . We found that the use of either the  $\tilde{\sigma}_{16}^B$  and  $\tilde{\sigma}_{16}^C$  or the  $\tilde{\sigma}_{1(16)1}^B$  and  $\tilde{\sigma}_{1(16)1}^C$  produced reasonable results: the  $\tilde{\sigma}_{16}$  parameters were slightly more accurate, but inclusion of both terms did not significantly improve the model. We also found that the parameter  $\tilde{\sigma}_{8(16)1}^C$  was not necessary to accurately reproduce the temperature dependence of dimethylsulfoxide. Table 11 shows the final parameters for S-containing compounds. See Figures 8–11 for comparison of different parameterizations of SM8T to experiment. Although the model accurately reproduces the temperature dependence of the free energy of solvation for thiols and sulfides, the model has an error twice as large for the thiophenes in the dataset. It is also apparent from Table 1 that the number of compounds in the S dataset is very small and that S-heterocycles and sulfoxides, in particular, are underrepresented and that the number of points per compound (particularly the thiophenes) is quite small. The user should use some caution when applying SM8T to unusual S-heterocycles and sulfoxides, particularly at extremes of the temperature range.

## 4. Concluding Remarks

We have developed the SM8T model to account for the temperature dependence of the aqueous free energy of solvation over the temperature range 273–373 K for compounds containing H, C, N, O, F, S, Cl, or Br. Note that the SM8T parameters involving only the atomic numbers of H, C, and O are taken to be the same as the previous SM6T parameters because SM6 and SM8 are almost the same for aqueous solutions; the other SM8T parameters are new in the article. The parametrization of this model involved the assembly of an extensive database of free energies of solvation at different temperatures with careful analysis of the resultant data and the selection of statistically meaningful parameters. The final SM8T model has a mean unsigned error of 0.08 kcal/mol for N-containing compounds; a mean unsigned error of 0.07 kcal/mol for compounds containing F, Cl, or Br; and a mean unsigned error of 0.07 kcal/mol for S-containing compounds. The model involves only six additional parameters for the *B* coefficient ( $\tilde{\sigma}_{171}^B$ ,  $\tilde{\sigma}_{761}^B$ ,  $\tilde{\sigma}_{762}^B$ ,  $\tilde{\sigma}_{b1}^B$ ,  $\tilde{\sigma}_{16}^B$ , and  $\tilde{\sigma}_{8(16)1}^B$ ) and two additional parameters for the *C* coefficient ( $\tilde{\sigma}_{c1}^C$  and  $\tilde{\sigma}_{16}^C$ ).

**Acknowledgment.** This work was supported by the U.S. Army Research Office under Multidisciplinary Research Program of the University Research Initiative (MURI) under Grant No. DAAD19-02-1-0176 and by the Office of Naval Research under Grant No. N00012-05-01-0538. Adam Chamberlin gratefully acknowledges support by an NIH Roadmap Fellowship on grant no. IT90DK070106-01 from the National Institute of Diabetes and Digestive and Kidney Diseases.

**Supporting Information Available:** Tables listing the individual compounds, the number of experimental points used for each compound, and the references for the experimental data are provided. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Denbigh, K. *The Principles of Chemical Equilibrium*, 4th ed.; Cambridge University Press: London, 1981.
- (2) Klotz, I. M.; Rosenberg, R. M. *Chemical Thermodynamics*, 5th ed.; Wiley: New York, 1994.
- (3) Cramer, C. J.; Truhlar, D. G. In *Free Energy Calculations in Rational Drug Design*; Reddy, M. R., Erion, M. D., Eds.; Kluwer: New York, 2001; p 63.
- (4) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **2003**, *119*, 1661.
- (5) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
- (6) Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertrán, J., Eds.; Kluwer: Boston, 1996; p 1.
- (7) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
- (8) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.
- (9) Grant, D. J. W.; Higuchi, T. *Solubility Behavior of Organic Compounds*; Wiley: New York, 1990.
- (10) Feig, M.; Brooks, C. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217.
- (11) Jorgensen, W. L. *Science* **2004**, *303*, 1813.
- (12) Raha, K.; Merz, K. M. Jr. *Annu. Rep. Comput. Chem.* **2005**, *1*, 113.
- (13) Carlsson, J.; Ander, M.; Nervall, M.; Aqvist, J. *J. Phys. Chem. B* **2006**, *110*, 12034.
- (14) Su, Y.; Gallicchio, E.; Das, K.; Arnold, E.; Levy, R. M. *J. Chem. Theory Comput.* **2007**, *3*, 256.
- (15) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.
- (16) Hawkins, G.; Cramer, C. J.; Truhlar, D. G. *J. Chim. Phys. Phys. Chim. Biol.* **1997**, *94*, 1448.
- (17) Berneche, S.; Nina, M.; Roux, B. *Biophys. J.* **1998**, *75*.
- (18) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401.
- (19) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187.
- (20) Roux, B.; Berneche, S.; Im, W. *Biochemistry* **2000**, *39*, 13295.
- (21) Im, W.; Roux, B. *J. Mol. Biol.* **2002**, *322*, 851.

- (22) Winget, P.; Cramer, C. J.; Truhlar, D. G. *Environ. Sci. Technol.* **2000**, *34*, 4733.
- (23) Hamad, S.; Lago, S.; Mejias, J. A. *J. Phys. Chem. A* **2002**, *106*, 9104.
- (24) Mirbagheri, S. A. *Int. J. Environ. Sci. Technol.* **2004**, *1*, 141.
- (25) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 133.
- (26) Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 5665.
- (27) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.
- (28) Marenich, A.; Olson, R.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2011.
- (29) McQuarrie, D. A.; Simon, J. D. *Physical Chemistry: A molecular Approach*; University Science Books: Sausalito, CA, 1997.
- (30) Atkins, P. *Physical Chemistry*, 6th ed.; Freeman: New York, 1998.
- (31) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Soln. Chem.* **1981**, *10*, 563.
- (32) Wick, C. D.; Siepmann, J. I.; Schure, M. R. *J. Phys. Chem. B* **2003**, *107*, 10623.
- (33) Staudinger, J.; Roberts, P. V. *Chemosphere* **2001**, *44*, 561.
- (34) Stephenson, R. M.; Malanowski, S. *Handbook of the Thermodynamics of Organic Compounds*; Elsevier: New York, 1987.
- (35) Yalkowsky, S. H.; He, Y. *Handbook of Aqueous Solubility Data*; CRC Press: New York, 2003.
- (36) Oe, S. *Computer Aided Data Book of Vapor-Pressure*; Data Book Publishing Co.: Tokyo, 1978.
- (37) Yaws, C. L. *Yaws' Handbook of Thermodynamic and Physical Properties of Chemical Compounds*; Knovel: Norwich, 2003.
- (38) Gautreaux, M. F.; Coates, J. *AIChE J.* **1955**, *1*, 496.
- (39) Gmehling, J.; Menke, J.; Schiller, M. *Activity Coefficients at Infinite Dilution: C<sub>10</sub>-C<sub>36</sub> with O<sub>2</sub>S and H<sub>2</sub>O*; Dechema: Frankfurt, 1994; Vol. 9, part 4.
- (40) Hovorka, S.; Dohnal, V.; Roux, A. H.; Roux-Desgranges, G. *Fluid Phase Equilib.* **2002**, *201*, 135.
- (41) Christie, A. O.; Crisp, D. J. *J. App. Chem.* **1967**, *17*.
- (42) Jones, F. M. I.; Arnett, E. M. *Prog. Phys. Org. Chem.* **1974**, *11*, 263.
- (43) Franks, F.; Watson, B. *Trans. Faraday Soc.* **1981**, *10*, 563.
- (44) Ben-Naim, A. *Solvation Thermodynamics*; Plenum: New York, 1978.
- (45) Winget, P.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2000**, *104*, 4726.
- (46) Lemmon, E. W.; Goodwin, A. R. H. *J. Phys. Chem. Ref. Data* **2000**, *29*, 1.
- (47) Poling, B. E.; Prausnitz, J. M.; O'Connell, J. P. *The Properties of Gases and Liquids*, 5th ed.; McGraw-Hill: New York, 2001.
- (48) Stephenson, R.; Stuart, J. *J. Chem. Eng. Data* **1986**, *31*, 56.
- (49) Chen, H.; Badenschier, S. M.; Shonnard, D. R. *Ind. Eng. Chem. Res.* **2002**, *41*, 4440.
- (50) Wang, Y.; Wong, P. K. *Sepu* **2003**, *21*, 614.
- (51) Strekowski, R. S.; George, C. J. *Chem. Eng. Data* **2005**, *50*, 804.
- (52) Lodge, K. B.; Danso, D. *Fluid Phase Equilib.* **2007**, *253*, 74.
- (53) Slocum, E. W.; Dodge, B. F. *AIChE J.* **1964**, *10*, 364.
- (54) Clarke, E. C. W.; Glew, D. N. *Trans. Faraday Soc.* **1966**, *62*, 539.
- (55) Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. G.; Winget, P.; Cramer, C. J.; Truhlar, D. G. *Minnesota Solvation Database 3.0*; University of Minnesota, Minneapolis, MN, 2007; the software may be obtained at <http://comp.chem.umn.edu/mnsol/>.
- (56) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305.
- (57) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8552.
- (58) Hawkins, G. D.; Zhu, T.; Li, J.; Chambers, C. C.; Giesen, D. J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. In *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M. A., Eds.; American Chemical Society: Washington, DC, 1998; p 201.
- (59) Cramer, C. J.; Truhlar, D. G. In *Trends and Perspectives in Modern Computational Science*; Maroulis, G.; Simos, T. E., Eds.; Brill Academic: Amsterdam, 2006; p 112.
- (60) Luque, F. J.; Bachs, M.; Aleman, C.; Orozco, M. *J. Comp. Chem.* **1996**, *17*, 806.
- (61) Hawkins, G.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 7147.
- (62) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; Wiley: New York, 2002.
- (63) Daudel, R. *Quantum Theory of Chemical Reactivity*; Reidel: Dordrecht, 1973.
- (64) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164.
- (65) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J. *Am. Chem. Soc.* **1990**, *112*, 6127.
- (66) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.
- (67) Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2046.
- (68) Marenich, A. V.; Olson, R. M.; Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2055.
- (69) Constanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, *65*, 1.
- (70) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Veillard, A., Eds.; Kluwer: Boston, 1981; Vol. 2, p 25.
- (71) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 9.
- (72) Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 1057.
- (73) Makhadadze, G. I.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307.
- (74) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133.
- (75) Shimizu, S.; Chan, H. S. *J. Chem. Phys.* **2000**, *113*, 4683.
- (76) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
- (77) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.
- (78) Hermann, R. B. *J. Phys. Chem.* **1975**, *79*, 163.
- (79) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comp. Chem.* **1995**, *16*, 422.
- (80) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
- (81) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6532.
- (82) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117.
- (83) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385.
- (84) Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1996**, *61*, 8720.
- (85) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 2061.
- (86) Giesen, D. J.; Hawkins, G.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1997**, *98*, 85.
- (87) Hawkins, G.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257.
- (88) Hawkins, G.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1998**, *63*, 4305.
- (89) Zhu, T.; Li, J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 5503.
- (90) Wohlfarth, C. In *CRC Handbook of Chemistry and Physics*; Lide, D. R., Ed.; CRC: Boca Raton, 1995; Vol. 75, section 6; p 189.
- (91) Chamberlin, A. C.; Kelly, C. P.; Xidos, J. D.; Li, J.; Thompson, J. D.; Hawkins, G. D.; Winget, P. D.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM*, 6.2; University of Minnesota, Minneapolis, MN, 2007; the software may be obtained at <http://comp.chem.umn.edu/mn-gsm>.
- (92) Chamberlin, A. C.; Kelly, C. P.; Pu, J.; Thompson, D. J.; Xidos, J. D.; Li, J.; Zhu, T.; Hawkins, G. D.; Chuang, Y.-Y.; Fast, P. L.; Lynch, B. J.; Liotard, D. A.; Rinaldi, D.; Gao, J.; Cramer, C. J.; Truhlar, D. G. *GAMESSPLUS*, 4.8; University of Minnesota, Minneapolis, 2006; based on the General Atomic and Molecular Electronic Structure System (GAMESS); the software may be obtained at <http://comp.chem.umn.edu/gamesplus>.
- (93) Chamberlin, A. C.; Kelly, C. P.; Thompson, D. J.; Lynch, B. J.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Zhu, T.; Volobuev, Y.; Dupuis, M.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *SMXGAUSS*, 3.4.1; University of Minnesota, Minneapolis, MN, 2006; the software may be obtained at <http://comp.chem.umn.edu/smxgauss>.
- (94) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nagatsuki, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.01; Gaussian: Pittsburgh, PA, 2003.
- (95) Nakamura, H.; Xidos, J. D.; Chamberlin, A. C.; Kelly, C. P.; Valero, R.; Thompson, D. J.; Li, J.; Hawkins, G. D.; Zhu, T.; Lynch, B. J.; Volobuev, Y.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *HONDOPLUS*, 5.1; University of Minnesota, Minneapolis, MN, 2007; based on HONDO v. 99.6; the software may be obtained at <http://comp.chem.umn.edu/hondoplus>.
- (96) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.
- (97) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

- (98) Hehre, W. J.; Radom, L.; Schleyer, P. von R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
- (99) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257.
- (100) Lasdon, L. S.; Waren, A. D.; Jain, A.; Ratner, M. *ACM Trans. Math. Software* **1978**, *4*, 34.
- (101) Lee, I.-J.; Jung, G.-S.; Kim, K. *J. Soln. Chem.* **1994**, *23*, 1283.
- (102) Bartmess, J. E. Negative Ion Energetics Data; In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD 20899, June 2005; <http://webbook.nist.gov> (accessed Oct. 2007).
- (103) Bartels, D. M.; Takahashi, K.; Cline, J. A.; Marin, T. W.; Jonah, C. D. *J. Phys. Chem. A* **2005**, *109*, 1299.
- (104) Hunter, E. P.; Lias, S. G. Proton Affinity Evaluation; In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD 20899, June 2005; <http://webbook.nist.gov> (accessed Oct. 2007).
- (105) NIST Computational Chemistry Comparison and Benchmark Database; In *NIST Standard Reference Database Number 101*; Johnson III, R. D., Ed.; National Institute of Standards and Technology: Gaithersburg, MD 20899, Aug 2005; Vol. Release 12; <http://srdata.nist.gov/cccbdb> (accessed Nov. 2007).