Parametrization of a Universal Solvation Model for Molecules Containing Silicon

Paul Winget, Jason D. Thompson, Christopher J. Cramer,* and Donald G. Truhlar*

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

Received: January 29, 2002

The SM5.42 solvation model is extended to include compounds containing Si. The new parameters are based on a data set of 13 octanol/water partition coefficients (which we convert into 13 differential free energies of solvation), three absolute solvation energies, and one pK_a. The data set includes compounds containing C, H, O, and Si. We carried out parametrizations using compounds in the data set that do not contain bonds between Si and O (i.e., eight differential free energies of solvation and three absolute free energies of solvation for nine compounds) at the HF/MIDI!, HF/MIDI!6D, HF/6-31G*, HF/6-31+G*, HF/cc-pVDZ, BPW91/MIDI!, BPW91/MIDI!6D, BPW91/DVZP, B3LYP/MIDI!, AM1, and PM3 levels of theory. The mean unsigned errors over the eight differential free energies of solvation and three absolute solvation energies for these levels of theory are in the range of 0.48−0.53 kcal/mol. We used five additional differential free energies of solvation for five compounds that do contain O−Si bonds to parametrize the BPW91/6-31G* level of theory. The resulting mean unsigned error over all 13 differential free energies of solvation and absolute free energies of solvation is 0.44 kcal/mol for this level of theory.

1. Introduction

The standard-state free energy of solvation, $\Delta G_S°$, is a fundamental quantity that describes the energy of interaction between a molecule and the solvent.[1−3] In our previous work,[4−9] we have shown that this interaction energy can be modeled as a sum of two components, a bulk electrostatic term and an additional term that depends on the solvent-accessible surface areas of the various kinds of atoms in the solute, with coefficients called atomic surface tensions. The most complete model of this type that we have advanced to date is called solvation model 5.42.[10−13] In this model, the atomic surface tensions are represented by functional forms denoted as version 5,[14−20] and the electrostatic terms are based on atomic class IV point charges[21] obtained by charge model 2.[22] The model was parametrized[10−12] for 275 solutes containing H, C, O, N, F, S, P, Cl, Br, and I against 2135 data points in 91 solvents. The present work extends the parametrization to Si on the basis of 16 data points in three solvents (water, hexadecane, and 1-octanol)

Section 2 is a review of the theory involved in the calculation of solvation free energies using SM5.42. Section 3 is a description of the experimental data. The parametrization methods and results are presented in section 4 and discussed in section 5.

2. Theory

Here, we summarize the essential features of the SM5.42 model. The standard-state free energy of solvation in this model is given by

$$\Delta G_S°(\mathrm{SM5.42}) = \Delta G_{\mathrm{ENP}} + G_{\mathrm{CDS}} \qquad (1)$$

where ΔG_{ENP} is the change in electronic and solute nuclear energy and solvent electric polarization energy when a solute is immersed in a continuous medium of dielectric constant ϵ and G_{CDS} is a cavitation−dispersion solvent-structure term that accounts for first-solvation-shell effects. The ΔG_{ENP} term accounts for bulk electrostatic effects, and the G_{CDS} term for SM5.42 is written as a sum over contributions from individual atoms k of the solute as follows:

$$G_{\mathrm{CDS}} = \sum_k \sum_j \sum_\delta A_k(\mathbf{R}) \sigma_{Z_k j \delta} f_{Z_k j}(\{Z_{k'}, R_{k'k''}\}) S_\delta \qquad (2)$$

In this expression, $A_k(\mathbf{R})$ is the solvent-accessible surface area of atom k, which depends on the complete three-dimensional geometry $\mathbf{R}$ of the solute; $\sigma_{Z_k j \delta}$ is an atomic surface tension coefficient that depends on the atomic number Z_k of atom k ($Z_k = 1, 6, 7, ...$ for H, C, N, ...), the geometrical factor index j, and the solvent descriptor index δ; $f_{Z_k j}(Z_{k'},\{R_{k'k''}\})$ is a geometrical factor that depends on the atom (k) and on the collection of all the atomic numbers $Z_{k'}$ and interatomic distances $R_{k'k''}$ in the molecule; and S_δ is a solvent descriptor. Because $A_k(\mathbf{R})$ is the solvent-accessible area[23] (rather than, for example, the van der Waals radius), it is necessary to specify the solute van der Waals radius and the solvent-shell half-thickness to calculate it. For SM5.42, all solute van der Waals radii are given by the values of Bondi[24] (for Si, this radius is 2.1 Å), and the solvent-shell half-thickness (also called solvent radius) is set equal to zero.

The calculation of the electrostatic term, ΔG_{ENP}, is fully described elsewhere.[11−13] It contains the electric polarization free energy, G_P, of the solvent and also the change in the electronic and nuclear energy of the solute. This term is computed by a quantum mechanical self-consistent reaction-field method,[11−13] and it depends on the intrinsic Coulomb radii ρ_Z of the various elements (with atomic numbers Z) and the solvent dielectric constant ϵ. The critical portion of the GB

---

* Corresponding authors. E-mail: cramer@chem.umn.edu, truhlar@umn.edu. Tel: 612-624-0859 (C.J.C.), 612-624-7555 (D.G.T.). Fax: 612-626-2006 (C.J.C.), 612-626-9390 (D.G.T.)

10.1021/jp020277g CCC: $22.00

formulation is the reaction-field contribution $\hat{V}^{\mathrm{RF}}_{\mu\nu}$ to the $\mu\nu$ matrix element of the Fock operator; this contribution is written

$$\hat{V}^{\mathrm{RF}}_{\mu\nu} = \frac{\partial G_{\mathrm{P}}}{\partial P_{\mu\nu}} \tag{3a}$$

where $P_{\mu\nu}$ is an element of the density matrix and

$$G_{\mathrm{P}} = -\sum_{k,k'}\left(1 - \frac{1}{\epsilon}\right) q_k(P_{\mu\nu}) q_{k'}(P_{\mu\nu}) \gamma_{kk'} \tag{3b}$$

where k and k' label atoms, μ and ν label basis functions, $q_{k'}$ is the CM2 partial atomic charge on atom k', and $\gamma_{kk'}$ is the generalized Born (GB) Coulomb integral between atoms k and k'. The GB Coulomb integral depends on the interatomic distance between k and k' and the effective Coulomb radius of atoms k and k'. The effective Coulomb radius is computed from the intrinsic Coulomb radius and the solute geometry;[4,25] this is called the dielectric-descreening algorithm.

The CDS term involves six descriptors for organic solvents[10−12] (although there are seven terms in the sum over δ because one of the descriptors, β, appears as β for S_3 and as β^2 for S_5). The six descriptors are

n	refractive index at the wavelength of the Na D line
α	Abraham's[25−28] hydrogen bond acidity parameters $\Sigma\alpha_2$
β	Abraham's[25−28] hydrogen bond basicity parameter $\Sigma\beta_2$
γ	macroscopic molecular surface tension in units of cal mol^{-1} Å^{-2}
ϕ^2	square of the fraction ϕ of nonhydrogenic solvent atoms that are aromatic carbon atoms (aromaticity)
ψ^2	square of the fraction ψ of nonhydrogenic solvent atoms that are F, Cl, or Br (electronegative halogenicity)

The second term of eq 1 can be further broken down into two separate components

$$G_{\mathrm{CDS}} = G^{[1]}_{\mathrm{CDS}} + G^{[2]}_{\mathrm{CDS}} \tag{4}$$

where

$$G^{[1]}_{\mathrm{CDS}} = \sum_{\delta=1}^{3} S_\delta \sum_k \sum_j \sigma^{[1]}_{Z_k j \delta} f_{Z_k j}(\{Z_{k'}, R_{k'k''}\}) A_k(\mathbf{R}) \tag{5}$$

and

$$G^{[2]}_{\mathrm{CDS}} = \sum_{\delta=4}^{7} S_\delta \sigma^{[2]}_\delta \sum_k A_k(\mathbf{R}) \tag{6}$$

In the sums, δ runs over n, α, and β for $G^{[1]}_{\mathrm{CDS}}$ and over γ, β^2, ϕ^2, and ψ^2 for $G^{[2]}_{\mathrm{CDS}}$, k runs over atoms in the solute, and j runs over the indices of the geometrical functions. All of the terms in $G^{[2]}_{\mathrm{CDS}}$ are independent of the atomic number. The major difference between the terms $G^{[1]}_{\mathrm{CDS}}$ and $G^{[2]}_{\mathrm{CDS}}$ is that the atomic surface tensions in the first term are functions of the atomic numbers of the atoms and are multiplied by the geometrical factors and the solvent-accessible surface areas of the individual atoms, whereas in the second term, surface tensions are multiplied by the entire solvent-accessible surface area of the molecule. When a new solute element with atomic number Z is to be added to the parametrization, the existing parameters are held fixed, and the only new parameters that need to be determined are the intrinsic Coulomb radii ρ_Z for G_{ENP} and the $\sigma^{[1]}_{Zj\delta}$ coefficients and any new geometrical functions $f_{Zj}(\{Z_{k'}, R_{k'k''}\})$ and $f_{Z'j}(\{Z_{k'}, R_{kk'}\})$ in $G^{[1]}_{\mathrm{CDS}}$.

TABLE 1: Values of Solvent Descriptors for Water, Hexadecane, and 1-Octanol

solvent	n	α	β	γ	ϵ	σ	ψ
n-hexadecane	1.4345[a]	0.00[b]	0.00[b]	38.93[a]	2.040[a]	0.00	0.00
1-octanol	1.4295[a]	0.37[c]	0.48[c]	39.01[a]	9.862[a]	0.00	0.00
water	1.3328[a]	0.82[b]	0.35[b]	71.99[a]	78.355[a]	0.00	0.00

[a] Ref 35. [b] Ref 23. [c] Ref 24.

In all of our previous parametrizations of SM5.42R and SM5.42 models, there were enough data for water as solvent so that it was treated separately from the organic solvents, and the solvent descriptor for water was arbitrarily set equal to unity. However, for silicon, there are data in only three solvents: water, hexadecane, and 1-octanol. With such a limited number of data, we need to treat all the solvents together. Therefore, in the present parametrization, the water surface tension for silicon is treated in the same manner as are those for 1-octanol and hexadecane, that is,

$$\sigma^{[1]}_{14,\mathrm{water}} = \sigma^{[1]}_{14,n} n + \sigma^{[1]}_{14,\alpha} \alpha + \sigma^{[1]}_{14,\beta} \beta + \sum_{\delta=4}^{7} \sigma^{[2]}_\delta S_\delta \tag{7}$$

where n, α, β, γ, ϕ, and ψ are the values for water. The values for all of the solvent descriptors used in this article are given in Table 1.

It was also necessary to add another geometrical function for oxygen (with $Z_k = 8$ and $\delta = 2$ in eq 5) to account for the different chemical environment of oxygen when it is bonded to silicon. This function was set equal to a geometry-dependent switching function called a cutoff tanh, or COT, function.[14] This COT is written as

$$T(R_{kk'} | \bar{R}_{kk'}, \Delta R_{kk'}) = \begin{cases} \exp\left[-\left(\dfrac{\Delta R_{kk'}}{\Delta R_{kk'} - R_{kk'} + \bar{R}_{kk'}}\right)\right], & R_{kk'} \le \bar{R}_{kk'} + \Delta R_{kk'} \\ 0 \text{ otherwise} \end{cases} \tag{8}$$

where $R_{kk'}$ is the distance between two atoms k and k', $\bar{R}_{kk'}$ is the midpoint of the switch, and the range over which the function switches is $2\Delta R_{kk'}$. The oxygen surface tension term is therefore modified by adding a term for oxygen atoms connected to silicon:

$$\sigma^{[1]}_{k,\alpha}\Big|_{k=\mathrm{O}} = ... + \sigma^{[1]}_{\mathrm{OSi},\alpha} \sum_{k'=\mathrm{Si}} T(R_{kk'} | \bar{R}_{\mathrm{OSi}}, \Delta R_{\mathrm{OSi}}) \tag{9}$$

Because there are different systematic errors in the electrostatics of each combination of electronic structure theory and basis set, different sets of surface tension coefficients are used for each of them to compensate empirically.

When the SM5.42 model is used to calculate a solvation quantity at a gas-phase geometry, the results are denoted 5.42R. For any other application (for example, a potential of mean force calculation as a function of solute geometry or a calculation at a geometry optimized in the liquid), the results are labeled SM5.42, which is the generic name for both SM5.42R and SM5.42.

3. Experimental Partitioning Data

We searched the MedChem[28] database for compounds that contained silicon and had solvent/solvent or solvent/air partition coefficients. From these compounds, we eliminated compounds with more than one significantly different conformation. (A

TABLE 2: Experimental Free Energies of Solvation (kcal/mol) for Tetramethylsilane

solvent	ΔG_s
n-hexadecane	−2.92
1-octanol	−1.79
water	3.04

TABLE 3: Experimental Octanol/Water Partition Coefficients (log $P_{o/w}$) and Transfer Free Energies ($\Delta\Delta G_{o/w}$, kcal/mol) for the Compounds in This Article

molecule	log $P_{o/w}$	$\Delta\Delta G_{o/w}$
dimethylpropylsilane	3.22	−4.39
propyltrimethylsilane	3.84	−5.24
butyldimethylsilane	3.57	−4.87
butyltrimethylsilane	4.2	−5.73
dimethylphenylsilane	3.99	−5.44
phenyltrimethylsilane	4.72	−6.44
benzyltrimethylsilane	4.13	−5.64
hexamethyldisiloxane	4.2	−5.73
octamethyltrisiloxane	4.8	−6.55
decamethyltetrasiloxane	5.4	−7.37
4-trimethylsilylphenol	3.84	−5.24
triethoxyphenylsilane	2.99	−4.08
diphenyldiethoxysilane	4.92	−6.71

TABLE 4: Values Used in the New Cutoff tanh (COT) Function

parameter	value (Å)
$\bar{R}_{OSi}$	2.1
ΔR	0.3

significantly different conformation is one that is expected to have a significantly different free energy of solvation.) This left a total of 14 compounds with 16 pieces of data. Tetramethylsilane has water/air, hexadecane/water, and octanol/water partition coefficients, but the other 13 compounds have only octanol/water partition coefficients. Only tetramethylsilane had more than one experimental measurement of the partition coefficient in the database; for this piece of data, we took the average of the two octanol/water partition coefficients, 3.85 and 3.24. The experimental free energy of solvation values of tetramethylsilane in *n*-hexadecane, 1-octanol, and water are listed in Table 2. Table 3 lists the experimental octanol/water partition coefficients for the other 13 compounds.

All solvation energies in this article are standard-state values for a concentration of 1 M in the gas phase or in any solvent, and for simplicity we consistently omit the standard-state superscript in the rest of the article.

4. Parametrization

Atomic Radii. There are two radii for each atom in the solute for the calculation of the solvation free energy using eqs 1−6. The first radius is the intrinsic Coulomb radius that is used in the generalized Born equation of the ΔG_{ENP} calculation. This is the starting point for the calculation of the effective Coulomb radius using the dielectric-descreening approximation. The other atomic radius is the van der Waals radius that is needed to calculate the solvent-accessible surface area for use in the calculation of G_{CDS}. In the present work, for Si, both the intrinsic Coulomb radius and the van der Waals radius were taken to be the value given by Bondi,[24] 2.1 Å.

COT Function. The values chosen for $\bar{R}_{OSi}$ and ΔR_{OSi} are given in Table 4. Because all of the silicon−oxygen bonds in our training set were in the range of 1.63−1.65 Å, the value for $\bar{R}_{OSi}$ was set to a distance of 2.1 Å so that the function begins decreasing when the distance is significantly greater than a typical silicon−oxygen bond, and because ΔR is 0.3 Å, it vanishes completely by the time the two atoms are 2.4 Å apart.

Geometries. The 14 compounds were optimized using HF[30]/MIDI!.[31,32] Silanes were optimized to minimize gauche interactions with the alkyl chain, which is in its most fully extended conformation. The siloxanes were also taken to be in their most extended form, with the possible orientations of the hydrogens on the *tert*-butyl groups not considered.

Because tetramethylsilane is the only solute with multiple data, its solvent-accessible surface areas are very important. At the HF/MIDI! geometry, we obtain $A_H = 96.6$ Å^2, $A_C = 34.3$ Å^2, and $A_{Si} = 14.8$ Å^2 for this compound. As in all SM5.42 calculations, these values are calculated with a zero solvent effective radius.

All SM5.42R calculations in this article employ HF/MIDI! gas-phase geometries, except in parts of Table 16 and all of Table 19 where we employ BPW91/6-31G* gas-phase geometries.

Semiempirical Surface Tension Coefficients. For parametrization purposes, the geometry in liquid solution is assumed to be the same as that in the gas phase. In the previous SM5.42 parametrizations, the free energy of solvation was directly used in the parametrization. However, for silicon there is only one compound with a known absolute free energy of solvation. If the absolute free energy of solvation ΔG_A into solvent A and the partition coefficient (log $P_{A/B}$) for transfer from solvent B to solvent A is known, the free energy of solvation ΔG_B in solvent B can be calculated using

$$\log P_{A/B} = -\frac{\Delta G_A - \Delta G_B}{2.303RT} \equiv -\frac{\Delta\Delta G_{A/B}}{2.303RT} \qquad (10)$$

Because of the lack of absolute solvation free energies, the *differences* in the free energy of solvation (i.e., the transfer free energy $\Delta\Delta G_{A/B}$) are directly used in the parametrization in addition to the absolute free energy of solvation data.

From eq 3, we note that the accuracy of the SM5.42 solvation model hinges on accurate partial atomic charges. We use CM2 partial atomic charges because they can reproduce experimentally measurable quantities such as dipole moments[22] and thus represent a realistic description of a molecule's charge distribution. The CM2 charge model is a parametrized model that was not, however, parametrized for compounds containing bonds between O and Si. Therefore, we closely examined electrostatic properties calculated from CM2 charges for a variety of compounds containing Si−O bonds, and we compared these results to more accurate calculations. As a result of this study, we concluded that the BPW91/6-31G* partial charges were reasonable for molecules containing Si−O bonds, and we therefore selected the BPW91/6-31G* level of theory for parametrization of the SM5.42 model for compounds containing Si−O bonds.

We therefore divided the data set into two subsets: (i) 11 data for 9 compounds that do not contain bonds between Si and O and (ii) 5 data for 5 compounds that do contain bonds between Si and O. Using subset i, we determined the surface tension coefficients, $\sigma_{14,\delta}^{[1]}$ (for $j = 1$ and $\delta = 1, 2, 3$; note also that $f_{14,1} \equiv 1$), by multilinear regression for all 12 levels of theory for which SM5.42 had previously been parametrized for compounds without Si (i.e., for HF/MIDI!, HF/MIDI!6D, HF/6-31G*, HF/6-31+G*, HF/cc-pVDZ, BPW91/MIDI!, BPW91/MIDI!6D, BPW91/6-31G*, BPW91/DVZP, B3LYP/MIDI!, AM1, and PM3). Holding these parameters fixed, we then used subset ii to determine $\sigma_{OSi,\alpha}^{[1]}$ by linear regression for the

TABLE 5: Parameters for the Seven Hartree−Fock (HF) Parametrizations[a]

			HF				
	AM1	PM3	MIDI!	MIDI!6D	6-31G*	6-31+G*	cc-pVDZ
$\sigma_{14,n}^{[1]}$	−118.5	−126.6	−156.8	−168.3	−160.1	−218.2	−148.1
$\sigma_{14,\alpha}^{[1]}$	173.7	152.1	102.5	94.2	107.0	91.8	123.7
$\sigma_{14,\beta}^{[1]}$	161.1	149.0	142.2	129.3	117.3	117.3	145.6
$\sigma_{14,\mathrm{water}}^{[1]\ b}$	68.6	35.2	−50.8	−78.4	−60.0	−148.4	−20.5

[a] Units of cal/Å^2. [b] Calculated as described in the text.

TABLE 6: Parameters for the Five Density Functional (DF) and Hybrid Hartree−Fock Density Functional (HF−DF) Parmetrizations[a]

	BPW91				B3LYP
	MIDI!	MIDI!6D	6-31G*	DZVP	MIDI!
$\sigma_{14,n}^{[1]}$	−161.2	−163.1	−160.4	−158.8	−163.4
$\sigma_{14,\alpha}^{[1]}$	103.3	102.2	103.7	108.9	101.9
$\sigma_{14,\beta}^{[1]}$	126.6	116.1	122.9	96.0	125.9
$\sigma_{\mathrm{OSi},\alpha}^{[1]}$			−188.3		
$\sigma_{14,\mathrm{water}}^{[1]\ b}$	−60.8	−68.7	−60.2	−62.1	−65.4
$\sigma_{\mathrm{OSi,water}}^{[1]\ b}$			−154.4		

[a] Units of cal/Å^2. [b] Calculated as described in the text.

BPW91/6-31G* level of theory. The results of the regressions are given in Tables 5 and 6. The mean signed error (MSE) and mean unsigned error (MUE) in $\Delta\Delta G_{A/B}$ are shown in Tables 7 and 8.

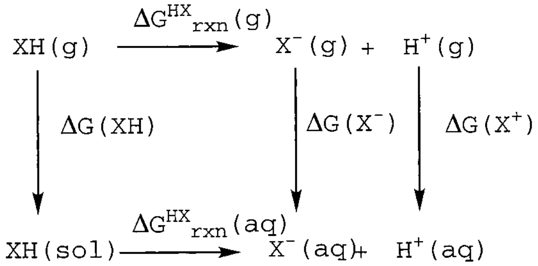
5. Discussion

Choice of Intrinsic Coulomb Radius. The choice of intrinsic Coulomb radii alters the relative contributions of the electrostatic terms and the first-solvation shell terms in the overall solvation free energy. However, we have previously found that reasonable changes in these radii have little effect on the overall accuracy of the predicted free energy of solvation. To test whether this holds true for silicon as well, we scanned a range of possible intrinsic Coulomb radii for one representative method, BPW91/6-31G*. We found that the choice of radius that was used had only a very small effect on the neutral compounds, with an MUE of 0.43−0.44 kcal/mol regardless of the choice of intrinsic Coulomb radius. The parameters and the MSE and MUE values for each choice of intrinsic Coulomb radius are given in Table 9.

One way to guide the choice for the intrinsic Coulomb radius is through analogy to other atoms comparable to silicon. The radii for several atoms in the SM5.42 model are shown in Table 10. Carbon, for instance, has an intrinsic Coulomb radius of 1.78 Å,[10−12] which is approximately 1.05 times larger than the van der Waals radius[24] of 1.7 Å. The intrinsic Coulomb radii for N and P are both significantly larger than their van der Waals radii, 1.92 and 2.27 Å, respectively. The values of intrinsic Coulomb radii for O and S, relative to van der Waals radii, are comparable to that of C; in particular, they are 0.08 and 0.18 Å greater than their van der Waals radii for intrinsic Coulomb radii of 1.60 and 1.98 Å, respectively. This result would suggest that a reasonable value for the intrinsic Coulomb radius is larger than the van der Waals radius.

Another guide to the choice of radius for silicon is the work of Aguilar et al.[33] For each basis set, they used a linear function of the atomic charge to assign the intrinsic Coulomb radius. In our work, the charge on silicon never varied to a large degree.

SCHEME 1



In the 14 neutral compounds in the parametrization set, the BPW91/6-31G* Löwdin charge on Si is in the range of 0.38 to 0.70, with an average of 0.50. The BPW91/6-31G* CM2 charges were slightly larger than this, in the range of 0.44−0.74. Using the linear function of Aguilar et al.,[33] the range of radii would be in the range of 2.63−2.81 Å. HF/MIDI! has CM2 charges that average 0.63 and range from 0.40 to 0.67. If these charges are used with the parameters that Aguilar et al. gave for STO-3G, the radius would be in the range of 2.27−2.34 Å.

For other atoms, although it was found[14] that the overall error in the free energy of solvation is relatively insensitive to the choice of intrinsic Coulomb radii (provided the atomic surface tensions are optimized for that choice of radius), the effect on ionic compounds is much greater. To see the effect of the Si intrinsic Coulomb radius on an ionic solvation property, we focused on orthosilicic acid, Si(OH)$_4$, whose gas-phase structure has been studied previously.[34] Orthosilicic acid has a known pK_a for the first proton (9.9),[35] and although an absolute calculation of the pK_a may be inaccurate, the difference in pK_a from another common acid could be used. To calculate this difference, we chose acetic acid with a pK_a of 4.76,[35] and we use Scheme 1. The ΔpK_a is then 5.14, which can be converted into a difference in the free energy of reaction of 7.0 kcal/mol. Only the lowest energy conformations at the HF/MIDI! level were used for both the neutral and anion. In the gas phase, the energy of reaction, $\Delta G_{\mathrm{rxn}}^{\mathrm{HX}}$, is the difference in Gibb's free energy of the anion and the neutral compound:

$$\Delta G_{\mathrm{rxn}}^{\mathrm{HX}} = G_{298}(\mathrm{X}^-) - G_{298}(\mathrm{XH}) \tag{11}$$

The difference in solvation free energies between the neutral ($\Delta G_S°$(XH)) and the anion ($\Delta G_S°(\mathrm{X}^-)$) of acid HX can be written as

$$\Delta\Delta G_{\mathrm{S}}^{\mathrm{HX}} = \Delta G_{\mathrm{S}}(\mathrm{X}^-) - \Delta G_{\mathrm{S}}(\mathrm{XH}) \tag{12}$$

Using eqs 11 and 12, the difference, $\Delta\Delta G_{\mathrm{rxn}}$(aq), in ΔG_{rxn}(aq) can be written

$$\Delta\Delta G_{\mathrm{rxn}}(\mathrm{aq}) = (\Delta G_{\mathrm{rxn}}^{\mathrm{Si(OH)_4}}(\mathrm{g}) + \Delta\Delta G_{\mathrm{S}}^{\mathrm{Si(OH)_4}}) - (\Delta G_{\mathrm{rxn}}^{\mathrm{C_2H_4O_2}}(\mathrm{g}) + \Delta\Delta G_{\mathrm{S}}^{\mathrm{C_2H_4O_2}}) \tag{13}$$

TABLE 7: MSE and MUE for the Seven Hartree−Fock (HF) Parametrizations

				HF				
	data	AM1	PM3	MIDI!	MIDI!6D	6-31G*	6-31+G*	cc-pVDZ
Silanes								
MSE[a]	10	0.14	0.13	0.13	0.14	0.14	0.12	0.13
MUE[b]	10	0.45	0.44	0.41	0.42	0.41	0.44	0.43
Multifunctional Si Compounds[c]								
MSE[a]	1	−1.36	−1.32	−1.24	−1.30	−1.28	−1.09	−1.29
MUE[b]	1	1.36	1.32	1.24	1.30	1.28	1.09	1.29
Overall								
MSE[a]	11	−0.02	−0.02	−0.02	−0.04	−0.06	−0.12	−0.04
MUE[b]	11	0.53	0.51	0.48	0.49	0.48	0.52	0.50

[a] Mean signed error of $\Delta\Delta G_{A/B}$ in kcal/mol. [b] Mean unsigned error of $\Delta\Delta G_{A/B}$ in kcal/mol. [c] Errors for 4-trimethylsilylphenol.

TABLE 8: MSE and MUE for the Five DF and HF-DF Parametrizations

		BPW91				B3LYP
	data	MIDI!	MIDI!6D	6-31G*	DZVP	MIDI!
Silanes						
MSE[a]	10	0.14	0.13	0.13	0.14	0.14
MUE[b]	10	0.43	0.42	0.41	0.43	0.42
Siloxanes						
MSE[a]	3			−0.16		
MUE[b]	3			0.44		
Multifunctional Si Compounds						
MSE[a]	3[c]	−1.34	−1.24	−0.47	−1.31	−1.31
MUE[b]	3[c]	1.34	1.24	0.52	1.31	1.31
Overall						
MSE[a]	16[c]	−0.02	−0.02	−0.03	−0.08	−0.01
MUE[b]	16[c]	0.50	0.48	0.44	0.50	0.48

[a] Mean signed error of $\Delta\Delta G_{A/B}$ in kcal/mol. [b] Mean unsigned error of $\Delta\Delta G_{A/B}$ in kcal/mol. [c] For BPW91/6-31G*, all three multifunctional Si compounds are considered. For all other levels of theory shown, only 4-trimethylsilylphenol is considered.

TABLE 9: Parameters, MSE, and MUE for a Range of Coulomb Radii (Å^2) for BPW91/6-31G*

Coulomb radii	1.8	2.1	2.4	2.8
BPW91/6-31G*				
$\sigma^{[1]}_{14,n}$[a]	−159.1	−160.4	−163.4	−167.2
$\sigma^{[1]}_{14,\alpha}$[a]	106.0	103.7	99.4	94.2
$\sigma^{[1]}_{14,\beta}$[a]	123.7	122.9	119.8	115.2
$\sigma^{[1]}_{\mathrm{OSi},\alpha}$[a]	−185.9	−188.3	−194.0	−200.5
$\sigma^{[1]}_{14,\mathrm{water}}$[a,b]	−56.2	−60.2	−68.6	−79.7
$\sigma^{[1]}_{\mathrm{OSi,water}}$[a,b]	−152.4	−154.4	−159.0	−164.4
MSE[c]	−0.03	−0.03	−0.03	−0.03
MUE[d]	0.43	0.44	0.44	0.44

[a] Units of cal/Å^2. [b] Calculated as described in the text. [c] Mean signed error of $\Delta\Delta G_{A/B}$ in kcal/mol. [d] Mean unsigned error of $\Delta\Delta G_{A/B}$ in kcal/mol.

TABLE 10: Intrinsic Coulomb and van der Waals Radii (Å) for Some Representative Elements

element	Coulomb	van der Waals
C	1.78	1.70
N	1.92	1.55
O	1.60	1.52
Si	2.10	2.10
P	2.27	1.80
S	1.98	1.80

TABLE 11: Solvation Energies (kcal/mol) for Acetic Acid and Its Anion Needed to Calculate the pK_a Difference

	BPW91/6-31G*	expt
$\Delta G_S°(C_2H_3O_2^-)$	−71.5	−77.0
$\Delta G_S°(C_2H_4O_2)$	−6.9	−6.7
$\Delta\Delta G_S^{C_2H_4O_2}$	−64.6	−70.3

TABLE 12: mPW1PW91/MG3S Energies (kcal/mol) for the Calculation of pK_a

	$\Delta G_{rxn}(g)$
acetic acid	346.6
orthosiliclic acid	358.7

TABLE 13: Solvation Free Energies (kcal/mol) of Orthosilicic Acid, Its Anion, and Differences in Their Solvation Free Energies over a Range of Intrinsic Coulomb Radii (Å) Using SM5.42R/BPW91/6-31G*

intrinsic Coulomb radius	1.8	2.1	2.4	2.8
$\Delta G_S°(Si(OH)_4)$	−25.4	−24.8	−23.6	−21.3
$\Delta G_S°(Si(OH)_3O^-)$	−89.1	−87.4	−84.6	−78.9
$\Delta\Delta G_S^{Si(OH)_4}$	−63.7	−62.6	−61.0	−57.6

As shown in Table 11, SM5.42R/BPW91/6-31G* gives a value of −64.6 kcal/mol for $\Delta\Delta G_S^{C_2H_4O_2}$, which is slightly lower than the experimental value of −70.3 kcal/mol.[36] Using mPW1PW91[37]/MG3S energies, which are shown in Table 12, for the gas-phase component of the reaction yields a target value of −69.7 kcal/mol for $\Delta\Delta G_S^{Si(OH)_4}$. (The MG3S basis is the "modified G3 semidiffuse" basis set, which is obtained from the MG3[38] or G3LargeMP2[39] basis set by deleting diffuse functions on H.) Table 13 shows that varying the intrinsic Coulomb radius from 1.8 to 2.8 Å causes the free energy of the anion to vary by 10.2 kcal/mol for BPW91/6-31G*, while that of the neutral varies by 4.1 kcal/mol. This variation leads to a change in $\Delta\Delta G_S^{Si(OH)_4}$ of 6.1 kcal/mol, with the largest difference in the solvation free energies between the neutral and anionic forms being −63.7 kcal/mol and an intrinsic Coulomb radius of 1.8 Å. Because usually it is best to make the intrinsic Coulomb radius larger than the van der Waals radius but the one datum we have that is most sensitive to the intrinsic Coulomb radius suggests a smaller value, we compromised by setting the intrinsic Coulomb radius equal to the van der Waals radius (i.e., a compromise of 2.1 Å was chosen for the Si intrinsic Coulomb radius).

Choice of Surface Tension Parameters. After the choice of intrinsic Coulomb radius is made, the surface tension coefficient parameters can be determined. Equation 4 allows for three new parameters for each new element as well as an additional three parameters for each geometrical factor introduced. Because there are a large number of possible parameter combinations to choose from, which surface tensions to fit is not obvious. To try to determine the best set of parameters, we carried out parametrizations with all the possible combinations

TABLE 14: Solvent-Accessible Surface Areas (Å^2) for the Compounds in the Training Set

solute	Si atoms	total Si area	area per Si atom
tetramethylsilane	1	14.8	14.8
dimethylpropylsilane	1	19.2	19.2
propyltrimethylsilane	1	13.9	13.9
butyldimethylsilane	1	19.2	19.2
butyltrimethylsilane	1	13.9	13.9
dimethylphenylsilane	1	18.8	18.8
phenyltrimethylsilane	1	13.5	13.5
benzyltrimethylsilane	1	13.8	13.8
hexamethyldisiloxane	2	31.9	15.9
octamethyltrisiloxane	3	51.8	17.3
decamethyltetrasiloxane	4	63.5	15.9
4-trimethylsilylphenol	1	13.6	13.6
triethoxyphenylsilane	1	12.9	12.9
diphenyldiethoxysilane	1	11.8	11.8

of $\sigma_{14,n}^{[1]}$, $\sigma_{14,\alpha}^{[1]}$, and $\sigma_{14,\beta}^{[1]}$ for HF/MIDI!. The MUE in $\Delta\Delta G_{A/B}$ dropped from 1.0 kcal/mol when no surface tension coefficients were used to 0.7 kcal/mol when only $\sigma_{14,n}^{[1]}$ was used. The reduction in the MUE by adding only $\sigma_{14,\alpha}^{[1]}$ or $\sigma_{14,\beta}^{[1]}$ was not as significant. Adding $\sigma_{14,\alpha}^{[1]}$ to $\sigma_{14,n}^{[1]}$ further reduced this error by 0.05 kcal/mol. Adding $\sigma_{14,\beta}^{[1]}$ to this combination reduced the error to 0.58 kcal/mol. Although the reduction in the overall error when adding the second and third parameters is not as large as when adding the first parameter, these parameters are still necessary. Despite indices of refraction (n) that are similar, *n*-hexadecane and 1-octanol are expected to have rather different first-solvation-shell effects. However, limiting the parametrization to just $\sigma_{14,n}^{[1]}$ would force the description to be nearly identical. Because the SM5.42 models are universal (i.e., they can be applied using *any* organic solvent), it is desirable to have these additional terms to better differentiate solvents despite the small effect on the overall MUE.

For our parametrization at the BPW91/6-31G* level of theory, we found that compounds that contained bonds between O and Si still had large errors without introducing a parameter specific to Si−O bonds. In the BPW91/6-31G* parametrization, the MUE over the siloxanes was 0.96 kcal/mol, which was significantly higher than that for the silanes, which had a MUE of 0.41 kcal/mol. We therefore tried fitting the $\sigma_{\mathrm{OSi}}^{[1]}$ surface tension (i.e., changing the O surface tension when it is near Si rather than changing the Si surface tension when it is near O). This approach is consistent with the rest of the elements in the model, where the surface tension of the more electronegative element is modified. Because acidic solvents can hydrogen bond with the oxygen atoms, the dependence of this surface tension coefficient on α was chosen. Another parametrization was also carried out using $\sigma_{\mathrm{HSi}}^{[1]}$; however, this process did not increase the accuracy of the calculated partition coefficients. The error after fitting these four surface tension parameters are in the same range as that for other chemical classes for the SM5.42 solvation model.

The magnitudes of the parameters, $\sigma_{14,n}^{[1]}$, $\sigma_{14,\alpha}^{[1]}$, $\sigma_{14,\beta}^{[1]}$, $\sigma_{\mathrm{OSi},\alpha}^{[1]}$, $\sigma_{14,\mathrm{water}}^{[1]}$, and $\sigma_{\mathrm{OSi,water}}^{[1]}$, are all around 200 cal/Å^2, which is in the same range as that of many of the parameters for other atoms. There is some variation in the parameters among the various combinations of the electronic structure level and basis set, which is of about the same magnitude as the variation in parameters for other elements.

Table 14 shows that the solvent-accessible area per silicon atom is relatively constant among all of the compounds. Because the solvent-accessible surface area per silicon atom is nearly the same, the resulting contribution due to G_{CDS} is similar per silicon atom. Using the BPW91/6-31G* parameters, G_{CDS} ranged over 7.5 kcal/mol in octanol, but the contribution per silicon atom was less than 0.9 kcal/mol.

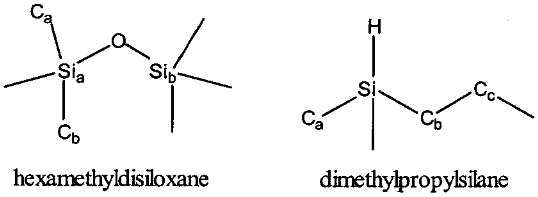


Figure 1. Two test molecules for optimization in the liquid phase.

TABLE 15: SM5.42/BPW91/6-31G* Geometrical Data for Two Test Molecules in the Gas Phase, 1-Octanol, and Water

	parameter[a]	gas	1-octanol	water
	Dimethylpropylsilane			
bond length (Å)	Si−H	1.508	1.514	1.514
bond length (Å)	C_a−Si	1.902	1.898	1.897
bond length (Å)	Si−C_b	1.908	1.909	1.907
bond length (Å)	C_b−C_c	1.544	1.543	1.543
bond angle (deg)	∠C_a−Si−H	108.3	108.1	108.1
	Hexamethyldisiloxane			
bond length (Å)	Si_a−O	1.670	1.684	1.692
bond length (Å)	Si_a−C_a	1.893	1.892	1.891
bond length (Å)	Si_a−C_b	1.897	1.894	1.893
bond angle (deg)	∠Si_a−O−Si_b	158.3	150.3	142.5
bond angle (deg)	∠C_a−Si_a−O	107.4	107.1	106.6
bond angle (deg)	∠O−Si_a−C_b	110.2	110.6	111.1

[a] As defined in Figure 1.

Geometry Optimization in Solution. Although the parametrization of the model was carried out using rigid gas-phase (HF/MIDI!) geometries, it is also possible to optimize the geometry in solution.[13] To test the effect of geometry optimization on these compounds, we considered two test molecules, dimethylpropylsilane and hexamethyldisiloxane, which are shown in Figure 1. Dimethylpropylsilane contains only nonpolar bonds: Si−H, C−H, and C−Si. Hexamethyldisiloxane contains a much more polar Si−O bond in addition to the types of bonds in dimethylpropylsilane. Optimizations in both 1-octanol and water were carried out using SM5.42/BPW91/6-31G*. The gas-phase geometries were optimized using BPW91[40,41]/6-31G*.[42−44] The geometrical data from these optimizations are summarized in Table 15.

For dimethylpropylsilane, there is only a small effect of carrying out the geometry optimization in solvent on the geometry. The silicon−carbon bonds are nearly constant, changing less than 0.005 Å from the gas phase to water. The carbon−carbon bonds change by 0.001 Å upon solvation in water. The largest change in bond length is between the silicon−hydrogen bonds, whose length increases from 1.508 to 1.514 Å in 1-octanol and water. There is little effect on the bond angles, which contract slightly.

For hexamethyldisiloxane, there is a much larger effect of liquid-phase optimization. The Si−O bond is lengthened when solvated, changing from 1.670 to 1.684 Å in 1-octanol and to 1.692 Å in water. The silicon−carbon bonds contract as they do in dimethylpropylsilane, changing from 1.897 to 1.893 Å in water. The Si−O−Si bond angle contracts drastically in solution, where the bond angle of 158.3° is reduced in 1-octanol to 150.3° and is even further reduced in water to 142.5°. The closing of this bond angle and the lengthening of the Si−O bond are also reflected in the O−Si−C bond angle that is slightly increased from 110.2° to 111.1° when this solute is solvated in water.

These geometrical changes are also reflected in the calculated partition coefficients shown in Table 16. There is only a small

TABLE 16: Free Energies of Solvation (kcal/mol) and Logarithms of Partition Coefficients Calculated Using SM5.42R/BPW91/6-31G*//HF/MIDI!, SM5.42R/BPW91/6-31G*, and SM5.42/BPW91/6-31G*

	SM5.42R/BPW91/6-31G*//HF/MIDI!	SM5.42R/BPW91/6-31G*	SM5.42/BPW91/6-31G*	expt
		Dimethylpropylsilane		
$\Delta G_{\text{octanol}}$	−3.11	−3.14	−3.16	
ΔG_{water}	1.79	1.82	1.79	
log $P_{\text{octanol/water}}$	3.59	3.63	3.63	3.22
		Hexamethyldisiloxane		
$\Delta G_{\text{octanol}}$	−4.50	−4.74	−4.85	
ΔG_{water}	0.60	0.42	−0.05	
log $P_{\text{octanol/water}}$	3.74	3.78	3.52	4.2

TABLE 17: Comparison of Results Using 6-31G* and Gordon's 6-31G* Basis for Silicon

			6-31G*			Si basis of ref 45[a]			expt
molecule	solvent 1	solvent 2	$\Delta G_S(1)$[b]	$\Delta G_S(2)$[c]	log $P_{1/2}$[d]	$\Delta G_S(1)$[b]	$\Delta G_S(2)$[c]	log $P_{1/2}$[d]	log $P_{1/2}$[d]
tetramethylsilane	water	air	2.51	0.00	−1.84	2.52	0.00	−1.85	−2.23
tetramethylsilane	hexadecane	air	−2.92	0.00	2.14	−2.91	0.00	2.13	2.14
tetramethylsilane	1-octanol	air	−1.26	0.00	0.92	−1.24	0.00	0.91	1.31
dimethylpropylsilane	1-octanol	water	−3.11	1.79	3.59	−3.11	1.78	3.58	3.22
propyltrimethylsilane	1-octanol	water	−2.38	2.81	3.80	−2.30	2.90	3.81	3.84
butyldimethylsilane	1-octanol	water	−3.63	1.98	4.11	−3.63	1.98	4.11	3.57
butyltrimethylsilane	1-octanol	water	−2.88	3.02	4.33	−2.79	3.12	4.33	4.2
dimethylphenylsilane	1-octanol	water	−5.64	−0.23	3.97	−5.63	−0.23	3.96	3.99
phenyltrimethylsilane	1-octanol	water	−5.16	0.52	4.16	−5.08	0.60	4.16	4.72
benzyltrimethylsilane	1-octanol	water	−5.72	0.71	4.71	−5.59	0.84	4.71	4.13
hexamethyldisiloxane	1-octanol	water	−4.50	0.60	3.74	−4.65	0.44	3.73	4.2
octamethyltrisiloxane	1-octanol	water	−7.59	−1.32	4.60	−7.84	−1.60	4.57	4.8
decamethyltetrasiloxane	1-octanol	water	−10.63	−2.85	5.70	−10.96	−3.23	5.67	5.4
4-trimethylsilylphenol	1-octanol	water	−8.53	−4.58	2.89	−8.47	−4.52	2.90	3.84
triethoxyphenylsilane	1-octanol	water	−12.60	−8.45	3.04	−12.71	−8.57	3.03	2.99
diphenyldiethoxysilane	1-octanol	water	−12.55	−6.03	4.78	−12.53	−6.01	4.78	4.92

[a] Using the 6-31G* basis set from ref 45 for silicon and the 6-31G* basis set from refs 42−44 for all the other elements. [b] The free energy of solvation, ΔG_S (kcal/mol), in solvent 1. [c] The free energy of solvation, ΔG_S (kcal/mol), in solvent 2. [d] The partition constant log $P_{\text{solvent 1/solvent 2}}$.

change in going from the HF/MIDI! to the BPW91/6-31G* gas-phase geometries because they are very similar. When optimizing in solution, the free energies of solvation increase more in water than they do in 1-octanol, which is reflected in a smaller partition coefficient. This effect is more pronounced in hexamethyldisiloxane, where optimization had a large effect on the geometry.

Variations on the 6-31G* Basis Set. A variation on the 6-31G* basis set[42−44] for silicon has been developed by Gordon.[45] To test the applicability of our parametrization when this change is made to the basis set, the partition coefficients for the training set were calculated using both variations of the 6-31G* basis with BPW91/SM5.42R. There is some variation in the overall free energy of solvation when changing the basis for Si, particularly in the siloxanes, which is shown in Table 17. This variation is generally small, less than 0.38 kcal/mol. For example, in hexamethyldisiloxane, the free energy of solvation in octanol changes from −4.50 using 6-31G* to −4.65 using Gordon's modification. This same favorable solvation effect is reflected in the free energy of solvation in water, where ΔG_S is reduced from 0.60 to 0.44 kcal/mol. Because nearly the same change occurs in both solvents, the partition coefficients are relatively constant, 3.74 and 3.73. The larger siloxanes show this same trend. The overall mean unsigned error for the two variations of 6-31G* remains nearly constant, as shown in Table 18.

Dependence of Solvation Free Energies on Solute and Solvent Structure. To illustrate the dependence on solute and solvent structure, the free energies of solvation calculated at the SM5.42R/BPW91/6-31G* and SM5.42/BPW91/6-31G* levels of theory for a small group of solutes and solvents are given in Table 19. The solvents selected are *n*-hexane, benzene, tetrahydrofuran (THF), 1,2-dichloroethane, *N*,*N*-dimethylacetamide (DMA), and water. These solvents exhibit a broad range of characteristics, indicated in Table 20. For each solvent, Table 20 lists the dielectric constant, ϵ, the index of refraction, n, Abraham's[26−28] hydrogen bond acidity parameter, α, Abraham's[26−28] hydrogen bond basicity parameter, β, the macroscopic surface tension, γ, the aromaticity factor, ϕ (i.e., the fraction of non-hydrogenic atoms that are aromatic carbon atoms), and the electronegative halogenicity factor, ψ (i.e., the fraction of non-hydrogenic atoms that are F, Cl, or Br).

TABLE 18: Mean Unsigned Error (MUE) in $\Delta\Delta G_{A/B}$ (kcal/mol) Using Two Variations of the 6-31G* Basis Set for Silicon, Broken Down by Functional Group

class	6-31G*	Si basis of ref 45[a]
silanes	0.41	0.41
siloxanes	0.44	0.44
multifunctional Si	0.52	0.51
all	0.44	0.44

[a] Using the 6-31G* basis set from ref 45 for silicon and the 6-31G* basis set from refs 42−44 for all the other elements.

TABLE 19: SM5.42R/BPW91/6-31G* and SM5.42/BPW91/6-31G* Free Energies of Solvation for Selected Solutes and Solvents

	n-hexane	benzene	THF	CH_2ClCH_2Cl	DMA[a]	water
			SM5.42R/BPW91/6-31G*			
$CH_3SiH_2CH_3$	−4.87	−4.89	−3.31	−4.42	−2.18	0.96
CH_3SiH_3	−5.45	−5.42	−3.56	−4.88	−2.29	0.62
SiH_3CH_2OH	−6.76	−7.31	−7.25	−8.18	−6.96	−5.99
			SM5.42/BPW91/6-31G*			
$CH_3SiH_2CH_3$	−4.88	−4.90	−3.33	−4.44	−2.21	0.93
CH_3SiH_3	−5.45	−5.42	−3.57	−4.89	−2.31	0.60
SiH_3CH_2OH	−6.78	−7.34	−7.31	−8.26	−7.04	−6.10

[a] *N*,*N*-Dimethylacetamide.

TABLE 20: Solvent Descriptors for Solvents in Table 19

	n-hexane	benzene	THF	CH_2ClCH_2Cl	DMA	water
ϵ	1.8819	2.2706	7.4257	10.1250	37.7810	78.355
n	1.3749	1.5011	1.4050	1.4448	1.4380	1.3328
α	0.00	0.00	0.00	0.10	0.00	0.82
β	0.00	0.14	0.48	0.11	0.78	0.35
γ	25.75	40.62	39.44	45.86	47.62	71.99
ϕ^2	0.00	1.00	0.00	0.00	0.00	0.00
ψ^2	0.00	0.00	0.00	0.25	0.00	0.00

The electrostatic contribution to the free energy of solvation depends, in part, on the dielectric constant of the solvent. Thus, as the dielectric constant increases from *n*-hexane to water, so does the electrostatic contribution to the free energy of solvation, particularly for the slightly polar molecule SiH_3CH_2OH.

The other solvent descriptors shown in Table 20, in conjunction with the corresponding surface tension parameters, $\sigma_{14,\delta}^{[1]}$ or $\sigma_{\delta}^{[2]}$, are used to model nonbulk electrostatic contributions to the free energy of solvation (i.e., first-solvation-shell effects). The index of refraction is related to the solvent's polarizability, which in turn is related to dispersion interactions of the solvent. Table 6 shows that $\sigma_{14,n}^{[1]}$ is -160.4 cal/Å^2 n. The acidity and basicity parameters, α and β, are related to the solvent's ability to donate and accept hydrogen bonds. Table 6 shows that the parameters, $\sigma_{14,\alpha}^{[1]}$ and $\sigma_{14,\beta}^{[1]}$, are 103.7 and 112.9 cal/Å^2, respectively, indicating that the nonbulk electrostatic part of the hydrogen bonding is unfavorable. The macroscopic surface tension, γ, of a particular solvent represents the energy required to make a surface in the solvent, a contribution to the free energy of solvation that is always unfavorable. The aromaticity and electronegative halogenicity factors are used to correct small systematic errors in treating aromatic solvents and solvents containing electronegative halogen atoms.[9] Note that the last three descriptors are multiplied by surface tension coefficients that are independent of atomic number (see the second term in eqs 4 and 6). Previously,[11] we have determined $\sigma_{\gamma}^{[2]}$, $\sigma_{\phi^2}^{[2]}$, and $\sigma_{\psi^2}^{[2]}$ to be equal to 0.33, -3.93, and -8.41 cal/Å^2, respectively, at the BPW91/6-31G* level of theory.

Table 19 shows how the different characteristics of each solvent affect the free energy of solvation of the three selected molecules, $CH_3SiH_2CH_3$, CH_3SiH_3, and SiH_3CH_2OH. For example, in *n*-hexane, a solvent with a similar index of refraction compared to that of the other solvents of interest and a small dielectric constant, the free energies of solvation for all three compounds are similar, with SiH_3CH_2OH affording more favorable electrostatic interactions with the solvent. Benzene has a larger index of refraction, dielectric constant, and hydrogen bond basicity than does *n*-hexane. The free energies of solvation of $CH_3SiH_2CH_3$ and CH_3SiH_3 are similar to their corresponding free energies of solvation in *n*-hexane, and this result can be attributed to a balance between the more favorable solute−solvent dispersion interactions in benzene (a consequence of benzene's larger index of refraction) and unfavorable interactions that correlate (at least according to our model) with benzene's hydrogen bond accepting ability. The molecule SiH_3CH_2OH has a more favorable free energy of solvation in benzene than in *n*-hexane, however, because of increased favorable electrostatic interactions.

For the solvent THF, the index of refraction is smaller than benzene's index of refraction and similar to the index of refraction of *n*-hexane. In addition, the hydrogen bond basicity parameter for THF is larger than that for either benzene or *n*-hexane. These two factors result in free energies of solvation of $CH_3SiH_2CH_3$ and CH_3SiH_3 that are approximately 1.0−1.5 kcal/mol less favorable than the corresponding free energies of solvation in *n*-hexane or in benzene. The molecule SiH_3CH_2OH affords more favorable electrostatic interactions in THF than in *n*-hexane and benzene; however, its free energy of solvation is similar to that in benzene because THF has a larger hydrogen bond basicity parameter than does benzene. In 1,2-dichloroethane, similar trends are observed for $CH_3SiH_2CH_3$ and CH_3SiH_3, but the free energy of solvation of SiH_3CH_2OH is the most favorable in this solvent. This result is most likely a combination of favorable solute−solvent dispersion interactions and relatively small hydrogen bond acidity and basicity parameters. For the solvent DMA, the free energies of solvation of $CH_3SiH_2CH_3$ and CH_3SiH_3 become even more unfavorable because of the larger β parameter for DMA. Table 19 shows that the free energy of solvation of SiH_3CH_2OH is not as favorable in DMA as it is in benzene, THF, or 1,2-dichloroethane, for the same reason.

For water, the free energies of solvation of $CH_3SiH_2CH_3$ and CH_3SiH_3 are significantly different than those in all of the other solvents because of the hydrophobic effect. For SiH_3CH_2OH, the hydroxyl group allows for favorable solute−solvent hydrogen-bonding interactions, but Table 20 shows that the free energy of solvation in water is the least favorable out of those of all the solvents listed, so these favorable interactions are outweighed by the hydrophobic contribution.

6. Conclusions

We have presented an extension to the SM5.42 solvation model for silicon. These parametrizations have attempted to balance the electrostatic and first-solvation-shell effects in a systematic manner. Because of the paucity of data that was available to us, this parametrization may be less robust than those for other elements, but the importance of adding the capability to handle silicon to the SM5.42 model motivates acceptance of this risk.

We found in this work that the mean errors in the free energies of solvation do not vary strongly as the level of electronic structure theory and the basis set are changed, which is very similar to the situation in previous parametrizations that we have carried out with class IV charges. Because the charge mapping yields similarly accurate charges for the various combinations of levels and basis sets, the electrostatic contributions to free energies of solvation are reasonably similar. Even to the extent that the electrostatic contributions do differ, the empirical atomic surface tensions systematically account for deviations between actual solvation free energies and the electrostatic contributions. Thus, the atomic surface tensions are the most important parameters of the theory, at least for neutral solutes. The Coulomb radii are the most important parameters for ions. Our parametrization strategy recognizes this duality and utilizes data for neutral and ionic solutes accordingly.

7. Availability of New Parameters

The new parameters for Si are available in several electronic structure programs, namely, AMSOL,[46] HONDO/S,[47,48] and MN-GSM.[49] The new Si parameters for the AM1 and PM3 Hamiltonians are available only in AMSOL, while the new parameters for all of the ab initio and DFT methods are available only in HONDO/S and MN-GSM. For technical reasons related to SCF convergence with nearly linear dependent basis functions, we recommend that the Si parameters for HF/6-31+G* be used only in the HONDO/S program.[50]

Acknowledgment. This work was supported in part by the National Science Foundation.

Supporting Information Available: Tables of solvation free energies for all of the compounds in each solvent and calculated partition coefficients. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Plenum: New York, 1992.

(2) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.

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

(4) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305.

(5) Cramer, C. J.; Truhlar, D. G. *Science (Washington, D.C.)* **1992**, *256*, 213.

(6) Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des*. **1992**, *6*, 629.

(7) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1995**, *16*, 442.

(8) Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 1057.

(9) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257.

(10) Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *288*, 293.

(11) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys*. **1998**, *109*, 9117.

(12) Li, J.; Zhu, T.; Hawkins, D. G.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 9.

(13) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys*. **1999**, *110*, 5503. Note that the SM5.42R model is a special case of the SM5.42 model in which the solute is "rigid" (R) (i.e., the solute is fixed at a gas-phase geometry). In SM5.42, in contrast, the calculations are based on geometries optimized in liquid solution or on molecular dynamics or reaction-path calculations in liquid solution. The SM5.42 and SM5.42R models are parametrized in rigid mode, but the same parameters are also used for calculations in nonrigid modes (i.e., without the R).

(14) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem*. **1996**, *100*, 16385.

(15) Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem*. **1996**, *61*, 8720.

(16) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 2061.

(17) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 7147.

(18) Giesen, D. J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc*. **1997**, *98*, 85.

(19) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257.

(20) Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem*. **1998**, *63*, 4305.

(21) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des*. **1995**, *9*, 87.

(22) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820.

(23) (a) Hermann, R. B. *J. Phys. Chem*. **1972**, *76*, 2754. (b) Lee, B.; Richards, F. M. *J. Mol. Biol*. **1971**, *55*, 379. (c) Richards, F. M. *Annu. Rev. Biophys. Bioeng*. **1977**, *6*, 151.

(24) Bondi, A. *J. Phys. Chem*. **1964**, *68*, 441.

(25) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc*. **1990**, *112*, 6127.

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

(27) (a) Abraham, M. H. *Chem. Soc. Rev*. **1993**, *22*, 73. (b) Abraham, M. H. *J. Phys. Org. Chem*. **1993**, *6*, 660.

(28) (a) Abraham, M. H. In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P., Murray, J. S., Eds.; Theoretical and Computational Chemistry 1; Elsevier: Amsterdam, 1994; pp 83−134. (b) Abraham, M. H.; Weatherby, P. K. *J. Pharm. Sci*. **1994**, *83*, 1087.

(29) Leo, A. J. *Pomona Medchem Database*; Biobyte Corporation: Claremont, CA, 1994.

(30) Roothaan, C. C. J. *Rev. Mod. Phys*. **1951**, *23*, 69.

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

(32) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1998**, *99*, 192.

(33) Aguilar, M. A.; Olivares del Valle, F. J. *Chem. Phys*. **1989**, *129*, 439.

(34) Pereira, J. C. G.; Catlow, C. R. A.; Price, G. D. *J. Phys. Chem. A* **1999**, *103*, 3252.

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

(36) Florian, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.

(37) Adamo, C.; Barone, V. *J. Chem. Phys*. **1998**, *108*, 664.

(38) Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. *Chem. Phys. Lett*. **1999**, *306*, 407.

(39) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703.

(40) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1998**, *38*, 3098.

(41) Perdew, J. P.; Burke, K.; Wang, K. *Phys. Rev. B: Condens. Matter* **1996**, *54*, 16533.

(42) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys*. **1972**, *56*, 2257.

(43) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

(44) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654.

(45) Gordon, M. S. *Chem. Phys. Lett*. **1980**, *76*, 163.

(46) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Winget, P.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *AMSOL*, version 6.7.2; University of Minnesota: Minneapolis, MN, 2002, based in part on *AMPAC*, version 2.1; Liotard, D. A.; Healy, E. F.; Ruiz, J. M.; Dewar, M. J. S.

(47) Xidos, J. D.; Thompson, J. D.; Li, J.; Hawkins, G. D.; Zhu, T.; Lynch, B. J.; Volobuev, Y.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *HONDO/S*, version 3.4.1; University of Minnesota: Minneapolis, MN, 2002, based on *HONDO*, version 99.6.

(48) Dupuis, M.; Marquez, A.; Davidson, E. R. *HONDO 99.6*, 1999, based on *HONDO 95.3*; Dupuis, M.; Marquez, A.; Davidson, E. R.; Quantum Chemistry Program Exchange (QCPE); Indiana University: Bloomington, IN, 47405.

(49) 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*, version 1.8; Univeristy of Minnesota: Minneapolis, MN, 2001.

(50) Further information about software is available at http://comp.chem.umn.edu/mccdir/software.html.