

## Density functional solvation model based on CM2 atomic charges

Tianhai Zhu, Jiabo Li, Gregory D. Hawkins, Christopher J. Cramer, and Donald G. Truhlar

Citation: *The Journal of Chemical Physics* **109**, 9117 (1998); doi: 10.1063/1.476521

View online: <http://dx.doi.org/10.1063/1.476521>

View Table of Contents: <http://aip.scitation.org/toc/jcp/109/20>

Published by the *American Institute of Physics*

---

---

**COMPLETELY**

**REDESIGNED!**



**PHYSICS  
TODAY**

*Physics Today* Buyer's Guide  
Search with a purpose.

## Density functional solvation model based on CM2 atomic charges

Tianhai Zhu, Jiabo Li, Gregory D. Hawkins, Christopher J. Cramer,  
and Donald G. Truhlar

*Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis,  
Minnesota 55455-0431*

(Received 22 December 1997; accepted 18 August 1998)

We extend the SM5 solvation model for calculating solvation free energies of a variety of organic solutes in both aqueous and organic solvents so that it can be employed in conjunction with high-level electronic structure calculations. The extension is illustrated by presenting three implementations based on density-functional theory (DFT). The three implementations are called SM5.42R/BPW91/MIDI!6D, SM5.42R/BPW91/DZVP, and SM5.42R/BPW91/6-31G\*. They have the following features: (1) They utilize gradient-corrected DFT with polarized double zeta basis sets to describe the electronic structure of a solute. The particular exchange-correlation functional adopted is Becke's exchange with the Perdew-Wang 1991 correlation functional, usually called BPW91. The MIDI!6D, DZVP, and 6-31G\* basis sets are used. (2) They employ fixed solute geometries in solvation calculations. The model is designed to predict solvation free energies based on any reasonably accurate gas-phase solute geometry. (3) The electric polarization in the solute-solvent system is described by the generalized Born approximation with self-consistent reaction-field solute partial atomic charges obtained from the CM2 class IV charge model. (4) The solvation effects within the first solvation shell are included in the form of SM5-type atomic surface tensions. Both DFT parameterizations are developed using 275 neutral solutes and 49 ions with gas-phase Hartree-Fock/MIDI! geometries. These solutes contain a wide variety of organic functional groups which include H, C, N, O, F, P, S, Cl, Br, and I atoms. For 2135 free energies of solvation of the neutral molecules in water and 90 organic solvents, SM5.42R/BPW91/MIDI!6D, SM5.42R/BPW91/DZVP, and SM5.42R/BPW91/6-31G\* yield mean unsigned errors in solvation free energies of 0.45 kcal/mol, 0.44 kcal/mol, and 0.43 kcal/mol, respectively. For 49 ions in water, SM5.42R/BPW91/MIDI!6D produces a mean unsigned error of 3.9 kcal/mol, while SM5.42R/BPW91/DZVP and SM5.42R/BPW91/6-31G\* give 3.6 kcal/mol and 3.9 kcal/mol, respectively.

© 1998 American Institute of Physics. [S0021-9606(98)50844-6]

### I. INTRODUCTION

In recent years, there has been a surge of interest in developing theoretical models for studying solvation effects because most chemical and biological processes occur in solution. Molecular models for solvent effects can be divided into two categories: the continuum-solvent approach,<sup>1</sup> which treats a solvent as a continuous dielectric medium, and the explicit-solvent approach,<sup>2</sup> which includes a large number of solvent molecules explicitly. While the explicit-solvent approach can provide microscopic details of solvation effects, it is computationally very demanding. As a consequence, a number of approximations have to be utilized to make such simulations feasible; for example, one typically resorts to a classical force field to describe the Hamiltonian for the system. Due to the large computational expense involved in running multiple simulations, it is often quite difficult to assess to what degree these approximations affect the overall simulation accuracy.<sup>2</sup> The alternative continuum-solvent approach does not yield atomic details for the solute-solvent interactions, but it eliminates sampling difficulties encountered in the explicit-solvent approach and is computationally much more efficient.<sup>1</sup> Continuum models are particularly

useful in calculating the absolute free energy of solvation, which is fundamentally important in solution thermochemistry.

The present paper is concerned with the SM $x$  continuum solvation models<sup>3-12</sup> and in particular with SM5 models.<sup>6-12</sup> Note that SM $x$  denotes Solvation Model  $x$  with  $x \geq 1$ , and Solvation Model 5 denotes a suite of Solvation Models with  $5 \leq x < 6$ , in particular SM5.0R, SM5.2PD, SM5.2R, SM5.4PD, SM5.4, and SM5.42R. (Note that PD denotes pairwise descreening and R denotes using rigid geometries during parameterization.) The SM5 suite of solvation models is unique in that several of these models (SM5.0R, SM5.42R, SM5.4, and SM5.42R) have been parameterized not only for aqueous solutions but also for general organic solvents. The common features in all SM $x$ -type solvation models prior to the present work ( $5 \leq x \leq 5.4$ ) are: (1) Semiempirical quantum-mechanical molecular-orbital theories such as Austin Model 1 (AM1), Parameterized Model 3 (PM3), the Modified Neglect of Diatomic Differential Overlap Model (MNDO), and the MNDO model with  $d$  orbitals (MNDO/ $d$ ) have been used to describe the electronic structure of a solute. (2) The solute-solvent electric polarization is described by the generalized Born (GB) approximation,<sup>13</sup> and the reac-

tion field is incorporated into the solvation calculations self-consistently.<sup>14</sup> (3) Empirical atomic-surface-tension functional forms have been developed to represent short-range solvation effects, which include the creation of a solute-sized cavity in the solvent, the solute-solvent dispersion interactions, and the rearrangement of solvent structure. (4) The atomic surface tensions are parameterized for the prediction of solvation free energies.<sup>3–12</sup> For example, in the recently constructed SM5.4 aqueous<sup>6</sup> and organic<sup>7,8,10</sup> solvation models, we have obtained mean unsigned errors (MUE) in solvation free energies of less than 0.5 kcal/mol for a variety of neutral solutes in both water and organic solvents. The general approach to atomic surface tensions employed in that work<sup>6–9</sup> is called the SM5 approach and will be used in this work as well.

Although previous SM $x$  solvation models have been very successful, improvements can still be envisioned with respect to two major aspects. The first is the replacement of semiempirical molecular-orbital theory by high-level electronic structure methods. These include *ab initio* Hartree–Fock molecular-orbital theory, Møller–Plesset perturbation theory, configuration interaction, and multiconfiguration self-consistent field theory. An alternative to these traditional molecular-orbital methods is density functional theory (DFT).<sup>15</sup> In recent years, DFT, particularly with the development of gradient-corrected exchange-correlation functionals, has established itself as an accurate and efficient first-principles approach.<sup>16–18</sup> Typically, the computational effort in traditional electronic structure methods that include electron correlation effects scales as  $N^5$  or higher, with  $N$  being the number of atoms in a system. This unfavorable scaling significantly hampers the applications of these methods to even moderately large molecular systems. The Kohn–Sham (KS) scheme<sup>15–18</sup> in DFT, on the other hand, formally scales as  $N^3$ . This cubic scaling continues to hold for DFT calculations that employ gradient-corrected functionals to treat exchange and correlation effects. This allows DFT to treat larger molecular systems than *ab initio* molecular-orbital methods can handle with comparable accuracy. Another attractive feature of DFT is that it is one of the most reliable theoretical tools for studying transition metal compounds. This makes DFT-based solvation models an obvious choice for theoretical investigations on, for instance, transition metal catalysis processes that take place in solution (e.g., metallocene catalysis of olefin polymerization<sup>19</sup>). Recently, a number of solvation models based on DFT have been proposed, and the results are very encouraging.<sup>20–25</sup> Based on these considerations, we here report a new solvation model that combines the power of the DFT approach to solute electronic structure and the SM5 approach to continuum solvation.

Another improvement we have made in the current solvation model is the use of the recently proposed Charge Model 2 (CM2).<sup>26</sup> As mentioned above, in the SM $x$  approach that we have developed the solvent polarization is computed on the basis of the GB method,<sup>13,14</sup> which is a very convenient approximation to the solution of the Poisson equation. The GB energy term depends upon the partial atomic charges in the solute, the shape of the solute cavity,

and the dielectric constant of the solvent. Consequently, the contribution of the polarization energy term to the reaction-field Fock operator is also a function of atomic charges. It is therefore critically important to have a charge model that can produce highly accurate atomic charges to best represent the electronic distribution of the solute. In our initial (unpublished) effort, we tried to use Mulliken<sup>27</sup> or Löwdin charges<sup>28</sup> to compute the polarization energy and the reaction-field Fock matrix in our DFT solvation model; we found that these charges were not stable with respect to varying the basis set. In addition, for a set of test molecules, the gas-phase dipole moments derived from these charges were found to have quite large errors compared to the experimental values. Following earlier work on the Charge Model 1 (CM1),<sup>29</sup> which was the first class IV atomic charge model, we recently constructed the CM2 charge model.<sup>26</sup> We found that the CM2 charges can reproduce charge-dependent observables such as dipole moments with high accuracy.<sup>26</sup> The CM2 charge scheme has therefore been incorporated into the DFT solvation model presented here.

Our current implementations of the DFT solvation model employ parameters that are determined using solute geometries that are fixed at gas-phase structures during parameterization. Solvation models based on parameterization with fixed solute geometries are called rigid models, or R models for short.<sup>12</sup> In principle, any reasonably accurate solute geometry can be used, but typically one starts with the best systematically available or computable estimate of the gas-phase geometry of the solute. We expect rigid models to be very useful in *ab initio* solvation calculations since geometry optimizations in the presence of solvent are often very time consuming, and for many (but not all) applications they provide little in the way of additional important information. By optimizing the parameters for gas-phase structures, the errors due to this constraint are minimized for typical applications.

The present solvation model is called SM5.42R to denote SM5-type atomic surface tensions, class IV charges based on CM2, and rigid geometries. The model employs the Becke–Perdew–Wang-1991 (BPW91) exchange-correlation functional<sup>30</sup> in the Kohn–Sham Hamiltonian. BPW91 is arguably the best theoretically justified gradient-corrected exchange-correlation functional available. It consists of Becke’s exchange functional and the Perdew–Wang 1991 correlation functional; the latter is called Generalized Gradient Approximation Two (GGA II) by the original authors. In one parameterization of the solvation model, the MIDI!6D basis set<sup>26</sup> was used; this basis set is identical to the original MIDI! basis set<sup>31</sup> except that the original five-function spherical harmonic  $d$  set is replaced by a six-function Cartesian  $d$  set. MIDI! and MIDI!6D are heteroatom-polarized split-valence basis sets, whose polarization functions were optimized to give accurate geometries and atomic charges.<sup>31</sup> Note that the naming of the MIDI basis by the original authors was intended to emphasize that it is MID-sized, between MINimal and MAXimal. The exclamation point, which is pronounced “bang,” denotes the addition of  $d$  basis functions on heteroatoms, which are all atoms except H and C. A second parameterization of the SM5.42R solvation

model uses the Double-Zeta-Valence Polarized (DZVP) basis set which is specifically designed for DFT calculations.<sup>32</sup> The DZVP basis set is also a polarized split-valence basis, but it includes six-function  $d$  sets on carbon as well as heteroatoms. A third parameterization is based on the standard 6-31G\* basis set,<sup>33–36</sup> which is a split-valence basis that again has six-function  $d$  sets on carbon and heteroatoms. (The six denotes that six contracted Gaussians are used as basis set for each core orbital, the 31G denotes that a contraction of three Gaussians plus a single Gaussian are used as basis functions for each valence orbital, and the\* denotes that  $d$  basis functions are added for atoms heavier than He.) Since the 6-31G\* basis set is not available for I, we used MIDI!6D instead for I atoms.

In semiempirical methods, computing high-level solute wave functions does not necessarily improve the accuracy of derived quantities such as free energies of solvation. “Errors” in the electrostatic terms are compensated by the parameterization of the atomic surface tensions; in fact it is a strength of the SMx approach that the intrinsic inseparability of electrostatics from surface terms does not lead to equally large uncertainties in their sum. A consequence, though, is that the parameters are method- and basis-set-dependent. The effort in developing the present DFT parameterizations is justified by the fact that DFT has a larger range of applicability and a higher accuracy than any other semiempirical methods; therefore it should be useful to couple it with a calibrated procedure for the evaluation of solvent effects. Thus we do not expect to obtain significantly more accurate solvation free energies than with our earlier models for solutes similar to those in the training set, but the results might be more reliable for molecules significantly different from those in the training set.

In the next section, we briefly describe the theoretical formalisms for the SM5.42R solvation model and the CM2 charge model. We then present details of the three parameterizations of the models in Sec. III. Section IV provides results and discussion, and Sec. V offers concluding remarks.

## II. THEORETICAL FORMALISM

In all SMx solvation models, the standard-state solvation free energy  $\Delta G_S^0$  of a solute molecule is computed by the following equation:<sup>1,3</sup>

$$\Delta G_S^0 = G_P + \Delta E_{EN} + G_{CDS}, \quad (1)$$

where  $G_P$  is the gain in electric polarization energy due to the mutual polarization of the solvent and solute,  $\Delta E_{EN}$  is the energy cost for distorting the solute to interact better with the polarized solvent, and  $G_{CDS}$  is the free energy of solvation due to cavitation, dispersion, and solvent structural changes (CDS) in the first solvation shell.  $G_{CDS}$  effectively also includes a correction for an systematic errors in  $G_P$ , to the extent that such errors may correlate with solvent-accessible surface area. In type-R models, the electronic relaxation part of  $\Delta E_{EN}$  is included explicitly, whereas the

nuclear part is absorbed implicitly in  $G_{CDS}$ . Thus  $\Delta E_{EN}$  becomes  $\Delta E_E$  in such models. The  $G_P$  term is calculated by the GB approximation:<sup>1,3,5,13,14,37,38</sup>

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

where  $q_k$  is the atomic charge on atom  $k$ ,  $\epsilon$  is the dielectric constant (relative permittivity) of the solvent, and  $\gamma_{kk'}$  is a Coulomb integral between atoms  $k$  and  $k'$ . This integral is computed using a modification<sup>6</sup> of a form originally suggested by Still and co-workers<sup>39</sup>

$$\gamma_{kk'} = \frac{1}{\sqrt{R_{kk'}^2 + \alpha_k \alpha_{k'} [\exp(-R_{kk'}/d_{kk'}) \alpha_k \alpha_{k'}]}}, \quad (3)$$

where  $R_{kk'}$  is the distance between two atoms  $k$  and  $k'$ ,  $d_{kk'}$  is a constant, and  $\alpha_k$  an effective radius which depends on a set of parameterized Coulomb radii.<sup>3,6</sup>

Equation (2) leads to the self-consistent reaction field (SCRf) Fock matrix with elements  $F_{\mu\nu}$  defined as<sup>1,3,6,14,37,38</sup>

$$F_{\mu\nu} = F_{\mu\nu}^0 + \frac{\partial G_P}{\partial P_{\mu\nu}}, \quad (4)$$

where  $\mu$  and  $\nu$  are basis function indices,  $F_{\mu\nu}^0$  is an element of the gas-phase Fock matrix, and  $P_{\mu\nu}$  is an element of the density matrix. In DFT, the gas-phase Fock operator is defined by the Kohn–Sham equation<sup>15</sup>

$$\hat{F}^0 \psi_i^0 = \left( -\frac{1}{2} \nabla^2 + V_C + V_{XC} \right) \psi_i^0 = \epsilon_i^0 \psi_i^0, \quad (5)$$

where  $\psi_i^0$  and  $\epsilon_i^0$  are the gas-phase one-electron orbital and orbital eigenvalue, respectively,  $V_C$  is the Coulomb potential, and  $V_{XC}$  is the exchange-correlation potential given by the functional derivative of the exchange-correlation energy  $E_{XC}[\rho]$  with respect to the electron density  $\rho$

$$V_{XC} = \frac{\delta E_{XC}[\rho]}{\delta \rho}. \quad (6)$$

In the SM5.42R solvation model, CM2 atomic charges are used to compute the  $G_P$  term in Eq. (2) and the solvated Fock matrix in Eq. (4) for SCRf solvation calculations. These atomic charges provide a simple way of describing the electrostatic interactions in a solute and those between a solute and a solvent. In the CM2 charge model,<sup>26</sup> the charge on atom  $k$  is calculated using the following expression:

$$q_k = q_k^L + q_k^C = q_k^L + \sum_{k' \neq k} B_{kk'} (D_{kk'} + C_{kk'} B_{kk'}), \quad (7)$$

where  $q_k$  is the CM2 charge on atom  $k$ ,  $q_k^L$  is the atomic charge obtained from Löwdin population analysis,<sup>28</sup>  $q_k^C$  is a correction term to the Löwdin charge  $q_k^L$ ,  $D_{kk'}$  and  $C_{kk'}$  are two empirical parameters to be determined, and  $B_{kk'}$  is Mayer's bond order<sup>40</sup> between atoms  $k$  and  $k'$ . The  $D_{kk'}$  and  $C_{kk'}$  parameters have been obtained for various combinations of theoretical methods and basis sets for chemical bonds that contain atoms with the atomic numbers of H, C,

N, O, F, Si, P, S, Cl, Br, and I.<sup>26</sup> At the BPW91/MIDI!6D, BPW91/DZVP, and BPW91/6-31G\* levels, CM2 yields a root-mean-square error of 0.20 D, 0.20 D, and 0.19 D, respectively, for the dipole moments of a set of 198 polar molecules. These errors are a factor of 6 better than the dipole moments computed from Mulliken population analysis, a factor of 3 better than the dipole moments computed by Löwdin population analysis, and even a factor of 2 better than the dipole moments computed from the continuous electron density.<sup>26</sup> Thus the CM2 mapping yields highly accurate partial atomic charges and makes the model more physical.

Based on the CM2 charge model, the contribution from the  $G_P$  term to the SCRF Fock matrix is

$$\frac{\partial G_P}{\partial P_{\mu\nu}} = - \left( 1 - \frac{1}{\epsilon} \right) \sum_k \sum_{k'} q_{k'} \gamma_{kk'} \left( \frac{\partial q_k^L}{\partial P_{\mu\nu}} + \frac{\partial q_k^C}{\partial P_{\mu\nu}} \right). \quad (8)$$

The derivative of  $q_k^L$  with respect to the density matrix is given by

$$\frac{\partial q_k^L}{\partial P_{\mu\nu}} = - \sum_{\lambda \in k} S_{\mu\lambda}^{1/2} S_{\lambda\nu}^{1/2}, \quad (9)$$

where  $\lambda$  is a basis function index, and  $S$  is the overlap matrix. The term  $\partial q_k^C / \partial P_{\mu\nu}$  is evaluated in Eqs. (10) and (11):

$$\frac{\partial q_k^C}{\partial P_{\mu\nu}} = \sum_{k' \neq k} \left[ \left( \frac{\partial B_{kk'}}{\partial P_{\mu\nu}} \right) (D_{kk'} + 2C_{kk'} B_{kk'}) \right], \quad (10)$$

$$\begin{aligned} \frac{\partial B_{kk'}}{\partial P_{\mu\nu}} = \frac{1}{2} \sum_{\lambda \in k} \sum_{\omega \in k'} [(\delta_{\lambda\mu} S_{\nu\omega} + \delta_{\lambda\nu} S_{\mu\omega})(PS)_{\omega\lambda} \\ + (\delta_{\omega\mu} S_{\nu\lambda} + \delta_{\omega\nu} S_{\mu\lambda})(PS)_{\lambda\omega}]. \end{aligned} \quad (11)$$

For each SCRF cycle, Eq. (8) is computed using Eqs. (9)–(11) to update the Fock matrix according to Eq. (4) until the solution-phase density matrix is converged.

The  $G_{\text{CDS}}$  term is computed as follows:<sup>3,6,7,12</sup>

$$G_{\text{CDS}} = \sum_k \sigma_k^A A_k(\{r_k, r_S^{(1)}\}) + \sigma^M \sum_k A_k(\{r_k, r_S^{(2)}\}), \quad (12)$$

where  $\sigma_k^A$  is the atomic-number-dependent contribution to the atomic surface tension of atom  $k$ , and  $A_k(\{r_k, r_S^{(i)}\})$ , with  $i=1$  or  $2$ , is the solvent-accessible surface area of atom  $k$  computed with a standard set of CDS solute radii  $\{r_k\}$  and one or another effective solvent radii,  $r_S^{(1)}$  and  $r_S^{(2)}$ . The CDS solute radii were set equal to the van der Waals radii of Bondi<sup>41</sup> without adjustments. In most previous SMx solvation models, the effective solvent radii were nonzero. But in a recent work,<sup>11</sup> we found that using zero effective solvent radii gives better accuracy in predicting solvation free energies of typical molecules. The best choice for the solvent radius for modeling various kinds of solvation phenomena is a subject for further study, but in the present work we decided to choose the value that gives the smallest root-mean-

square errors in solvation free energies of typical molecules. Therefore the effective solvent radii are chosen as zero in our DFT solvation model.

To compute the  $\sigma_k$  term in Eq. (12), we have previously introduced various empirical surface tension functional forms.<sup>3–12</sup> The present solvation model adopts the improved SM5 surface tension functional forms,<sup>12</sup> which are refinements of the original SM5-type surface tensions<sup>6</sup> and only depend on the geometry of the solute plus a list of solvent-specific descriptors. In SM5 models, the atomic surface tension of atom  $k$ ,  $\sigma_k^A$ , is a function of a geometry-dependent switching function called a cutoff tanh or a COT<sup>6,7,12</sup> and a set of underlying parameters called surface tension coefficients.<sup>5,6</sup> A COT is expressed as

$$\begin{aligned} T(R_{kk'} | \bar{R}, \Delta R) \\ = \begin{cases} \exp \left[ - \left( \frac{\Delta R}{\Delta R - R_{kk'} + \bar{R}} \right) \right] & R_{kk'} \leq \bar{R} + \Delta R \\ 0 & \text{otherwise} \end{cases}, \end{aligned} \quad (13)$$

where  $R_{kk'}$  is the distance between two atoms  $k$  and  $k'$  and  $\bar{R}$  is the midpoint of the switch while  $2\Delta R$  is the range over which the function switches. This function has the property that it vanishes identically for all  $R_{kk'}$  greater than  $\bar{R} + \Delta R$ , but it is continuous and has an infinite number of continuous derivatives for all  $R_{kk'}$ . The functional forms of the atomic surface tensions are the same as those in the SM5.2R models,<sup>12</sup> and they are given by:

$$\begin{aligned} \sigma_k^A |_{k=\text{H}} = \tilde{\sigma}_{\text{H}} + \sum_{k'=\text{C,O,N,S}} \left\{ T(R_{kk'} | \bar{R}_{\text{H}k'}, W) \right. \\ \times \left[ \tilde{\sigma}_{\text{H}k'} + \tilde{\sigma}_{\text{HO}}^{(2)} \sum_{\substack{k''=\text{H} \\ k'' \neq k \\ k'=O}} T(R_{k'k''} | \bar{R}_{\text{OH}}, W) \right. \\ \left. \left. + \tilde{\sigma}_{\text{HN}}^{(2)} \sum_{\substack{k''=\text{N} \\ k'=N \\ k' \neq k''}} T(R_{k'k''} | \bar{R}_{\text{NN}}, W) \right] \right\}, \end{aligned} \quad (14)$$

$$\begin{aligned} \sigma_k^A |_{k=\text{C}} = \tilde{\sigma}_{\text{C}} + \tilde{\sigma}_{\text{CC}} \sum_{\substack{k'=\text{C} \\ k' \neq k}} T(R_{kk'} | \bar{R}_{\text{CC}}, W) \\ + \tilde{\sigma}_{\text{CO}}^{(2)} \sum_{\substack{k'=C \\ k' \neq k}} T(R_{kk'} | \bar{R}_{\text{CC}}^{(2)}, W_{\text{CC}}) \\ + \tilde{\sigma}_{\text{CN}} \left[ \sum_{k'=N} T(R_{kk'} | \bar{R}_{\text{CN}}, W) \right]^2, \end{aligned} \quad (15)$$

$$\begin{aligned} \sigma_k^A|_{k=O} &= \tilde{\sigma}_O + \tilde{\sigma}_{OC} \sum_{k'=C} T(R_{kk'}|\bar{R}_{OC}^{(2)}, W_{OC}) \\ &\quad + \tilde{\sigma}_{OO} T\left(-\sum_{\substack{k'=O \\ k' \neq k}} T(R_{kk'}|\bar{R}_{OO}, W)\right) \bigg| R_{TT}, W_{TT} \\ &\quad + \tilde{\sigma}_{ON} \sum_{k'=N} T(R_{kk'}|\bar{R}_{ON}, W) \\ &\quad + \tilde{\sigma}_{OP} \sum_{k'=P} T(R_{kk'}|\bar{R}_{OP}, W), \end{aligned} \quad (16)$$

$$\begin{aligned} \sigma_k^A|_{k=N} &= \tilde{\sigma}_N + \tilde{\sigma}_{NC} \left\{ \sum_{k'=C} T(R_{kk'}|\bar{R}_{CN}, W) \right. \\ &\quad \times \left[ \sum_{\substack{k'' \neq k \\ k'' \neq k'}} T(R_{k'k''}|\bar{R}_{Ck''}, W) \right]^2 \bigg\}^{1.3} \\ &\quad + \tilde{\sigma}_{NC}^{(2)} \sum_{k'=C} \left[ T(R_{kk'}|\bar{R}_{CN}, W) \right. \\ &\quad \times \left. \sum_{k''=O} T(R_{k'k''}|\bar{R}_{CO}, W) \right] \\ &\quad + \tilde{\sigma}_{NC}^{(3)} \sum_{k'=C} T(R_{kk'}|\bar{R}_{NC}^{(2)}, W_{NC}), \end{aligned} \quad (17)$$

$$\sigma_k^A|_{k=F} = \tilde{\sigma}_F, \quad (18)$$

$$\sigma_k^A|_{k=P} = \tilde{\sigma}_P, \quad (19)$$

$$\begin{aligned} \sigma_k^A|_{k=S} &= \tilde{\sigma}_S + \tilde{\sigma}_{SS} \sum_{\substack{k'=S \\ k' \neq k}} T(R_{kk'}|\bar{R}_{SS}, W) \\ &\quad + \tilde{\sigma}_{SP} \sum_{k'=P} T(R_{kk'}|\bar{R}_{SP}, W), \end{aligned} \quad (20)$$

$$\sigma_k^A|_{k=Cl} = \tilde{\sigma}_{Cl}, \quad (21)$$

$$\sigma_k^A|_{k=Br} = \tilde{\sigma}_{Br}, \quad (22)$$

and

$$\sigma_k^A|_{k=I} = \tilde{\sigma}_I, \quad (23)$$

where  $\tilde{\sigma}_k$  and  $\tilde{\sigma}_{kk'}$  are atomic surface tension coefficients for atom  $k$  or for atom pair  $kk'$ , and the combination of COTs multiplied by a given surface tension coefficient is the surface tension functional associated with that coefficient. Note that<sup>12</sup>  $\bar{R}_{kk'} \equiv \bar{R}_{k'k}$  and that  $\bar{R}_{kx}$  is shorthand for  $\bar{R}_{kk'}|_{k'=x}$ . The surface tension coefficients in Eqs. (14)–(23) are either constants (for water) or depend on solvent properties (for organic solvents). The solvent properties that we use are the index of refraction,<sup>42</sup> the hydrogen-bond acidity and basicity,<sup>43–45</sup> the macroscopic surface tension,<sup>42</sup> and the dielectric constant.<sup>42</sup> The general form for  $\tilde{\sigma}_k$  or  $\tilde{\sigma}_{kk'}$  is<sup>7,12</sup>

$$\tilde{\sigma}_k = \hat{\sigma}_k^{(n)} n + \hat{\sigma}_k^{(\alpha)} \alpha + \hat{\sigma}_k^{(\beta)} \beta, \quad (24)$$

where  $n$  is the index of refraction, and  $\alpha$  and  $\beta$  are the hydrogen-bond acidity and basicity descriptors respectively. The  $\sigma^M$  term in Eq. (12) is computed as<sup>12</sup>

$$\sigma^M = \hat{\sigma}_{CS}^{(\gamma)} \gamma + \hat{\sigma}_{CS}^{(\beta^2)} \beta^2 + \hat{\sigma}_{CS}^{(\phi^2)} \phi^2 + \hat{\sigma}_{CS}^{(\psi^2)} \psi^2, \quad (25)$$

where  $\gamma$  is the macroscopic surface tension,  $\beta^2$  denotes the square of the hydrogen-bond basicity descriptor,  $\phi^2$  is the square of the fraction of the nonhydrogenic solvent atoms that are aromatic carbon atoms, and  $\psi^2$  is the square of the fraction of the nonhydrogenic solvent atoms that are electronegative halogen atoms. In other words,  $\phi$  is defined as the number of aromatic carbon atoms in the solvent divided by the total number of nonhydrogenic atoms in the solvent, and  $\psi$  is defined as the number of F, Cl, and Br atoms in the solvent divided by the total number of nonhydrogenic atoms in the solvent.

### III. PARAMETERIZATION

The parameterization procedure used in developing our DFT models is essentially the same as for the SM5.2R model.<sup>12</sup> In the following, we give a brief description of the major steps involved. Other details can be found in Ref. 12.

#### III.A. Training set

We have constructed several training sets for parameterization of previous SMx solvation models. The more recent ones are the SM5.4-aqueous model neutral and ion aqueous training sets,<sup>6</sup> the SM5.0R-aqueous model neutral training set,<sup>11</sup> the SM5.4-organic solvation model set,<sup>7,8,10</sup> the SM5.4-benzene and toluene solvation training sets,<sup>8</sup> the SM5.4-chloroform solvation training set,<sup>10</sup> the SM5.2R-aqueous model neutral and ion training set, and SM5.2R organic-solvent neutral training set.<sup>12</sup> The latter was obtained by merging and correcting the previous training sets for neutral nonphosphorus compounds, adding nine ions to the SM5.4 aqueous ion training set, and adding a new training set for phosphorus-containing compounds.<sup>12</sup> The references for the experimental data<sup>46–57</sup> are discussed elsewhere.<sup>6–8,10,12</sup> This merged training set is used here without any changes. For nonphosphorus containing compounds, it has 260 neutral solutes and 43 ions, composed of H, C, N, O, F, S, Cl, Br, and I atoms. For these solutes there are a total of 2084 neutral solvation free energies in 91 solvents (water plus 90 organic solvents), and 43 ionic solvation free energies, all in water. For phosphorus-containing compounds, the training set has 15 neutral solutes and 6 monocharged ions with a total of 57 experimental solvation free energies in water and 9 organic solvents.<sup>12</sup> Altogether there are 275 neutral solutes and 49 ions, there are 2135 free energies of solvation of neutrals in 91 solvents, and there are 49 free energies of solvation of ions in water.

TABLE I. Atomic radii ( $\text{\AA}$ ) used in calculating the electrostatic ( $\rho_k$ ) and nonelectrostatic ( $r_k$ ) contributions to the free energy of solvation in the SM5.42R solvation model.

$k$	$\rho_k^a$	$r_k^b$
H	0.91 <sup>c</sup>	1.20
C	1.78	1.70
N	1.92	1.55
O	1.60	1.52
P	2.27 <sup>d</sup>	1.80
S	1.98 <sup>d</sup>	1.80
F	1.50	1.47
Cl	2.13	1.75
Br	2.31	1.85
I	2.66	1.98

<sup>a</sup>Except where indicated, these values are from Chambers *et al.* (Ref. 6).<sup>b</sup>From Bondi (Ref. 41).<sup>c</sup>From Hawkins *et al.* (Ref. 9).<sup>d</sup>Optimized in this work.

### III.B. Solute Hamiltonians and geometries

Like the SM5.2R solvation models, the present DFT solvation model is designed to be used with any accurate solute geometry. In the parameterization, the gas-phase geometries of all 275 neutral solutes and 49 ions are optimized at the Hartree–Fock (HF) level using the MIDI! basis set (i.e., the original version of this basis with five  $d$  functions per  $d$  set).

### III.C. Code implementation

The SM5.42R solvation model has been incorporated into the DGAUSS density-functional module of the UNICHEM 4.0 package.<sup>58,59</sup> This involved creating a module in DGAUSS to compute the CM2 charges and modifying the program to construct the SCRF Fock matrix based on Eq. (4) and Eqs. (8)–(11). Additionally, the code used to calculate the  $G_{\text{CDs}}$  term as outlined in Eqs. (12)–(25) was incorporated by borrowing it from the AMSOL program.<sup>60</sup> In both the BPW91/MIDI!6D and BPW91/DZVP parameterizations, the A1 auxiliary basis set<sup>32,59</sup> in DGAUSS is used. In the BPW91/6-31G\* parameterization, the P1 auxiliary basis set<sup>59</sup> is used. We have checked these calculations by switching between the P1 or A1 auxiliary basis sets and found very little change in our results.

The present model has also been implemented in GAUSSIAN 94,<sup>61,62</sup> which yields essentially identical results. The fact that we obtained the same results with DGAUSS and GAUSSIAN 94 is a good check.

### III.D. Parameterization for water

The SM5.42R model presented in this paper was initially parameterized for water. There are two reasons for this. First, the variety of solutes for which solvation energies are available is much larger for water than for any other solvent.

TABLE II. Surface tension coefficients ( $\text{cal mol}^{-1} \text{\AA}^{-2}$ ) that depend on atomic numbers.

$k$	SM5.42R/BPW91/MIDI!6D				SM5.42R/BPW91/DZVP				SM5.42R/BPW91/6-31G*			
	$\hat{\sigma}_k^{(n)}$	$\hat{\sigma}_k^{(\alpha)}$	$\hat{\sigma}_k^{(\beta)}$	$\hat{\sigma}_k^{(\text{water})}$	$\hat{\sigma}_k^{(n)}$	$\hat{\sigma}_k^{(\alpha)}$	$\hat{\sigma}_k^{(\beta)}$	$\hat{\sigma}_k^{(\text{water})}$	$\hat{\sigma}_k^{(n)}$	$\hat{\sigma}_k^{(\alpha)}$	$\hat{\sigma}_k^{(\beta)}$	$\hat{\sigma}_k^{(\text{water})}$
H	40.47			99.12	38.61			98.82	39.90			99.02
C	73.99	29.01	1.21	113.43	74.34	32.29	12.39	120.67	67.40	26.88	−0.97	106.00
N	13.17	−94.13	97.40	−52.78	0.84	−120.39	74.12	−75.20	−2.88	−99.64	104.15	−90.08
O	−23.78	−31.51		−156.28	−44.86	−55.78		−189.36	−36.87	−28.64		−182.86
F	4.53			48.14	−2.14			36.32	−0.40			40.10
P	−376.17			−80.50	−252.56			−73.64	−441.54			−64.81
S	−77.90	−62.24	56.12	−87.77	−82.37	−72.96	56.79	−96.61	−75.70	−66.38	59.64	−79.43
Cl	−35.55			−4.45	−37.02			−6.16	−34.92			−1.84
Br	−47.18			−20.36	−48.51			−22.55	−46.98			−37.82
I	−50.10			−26.08	−52.81			−24.45	−51.23			−23.63
H, C	−109.83			−143.20	−111.17			−150.93	−107.49			−141.39
H, N	−87.46		−150.83	−243.02	−82.56		−122.58	−237.16	−77.64		−150.88	−218.15
H, N (2)	−191.08			−214.32	−188.51			−204.56	−188.54			199.86
H, O	−61.97	−316.97	−354.52	−352.55	19.40	−210.63	−285.18	−186.12	3.88	−267.35	−302.67	−203.96
H, O (2)	193.33			422.53	195.97			454.51	199.60			139.93
H, S	56.24			50.67	69.58			87.71	29.41			−6.91
C, C	−77.19			−78.03	−73.39			−68.88	−74.50			−75.73
C, C (2)	−13.07			−25.50	−16.88			−32.99	−7.77			−19.03
O, C	42.72		37.11	187.02	66.87		38.64	218.42	61.92		54.09	237.62
O, O	−49.96	112.39	−13.71	18.72	−42.21	120.87	−37.64	30.29	−46.54	103.71	−41.07	14.50
C, N	−79.73	106.32		16.70	−70.45	185.06		51.01	−63.67	114.92		48.88
N, C	−13.09	−59.98	−2.17	−65.02	−13.54	−68.06	−1.30	−67.29	−14.47	−60.78	−1.96	−67.54
N, C (2)				−308.61				−255.61				−288.29
N, C (3)	41.64			106.45	23.89			51.27	58.05			152.90
O, N	52.05	24.40	23.96	249.28	88.76	96.38	70.72	338.15	75.08	24.62	63.64	312.10
P, O	119.27			148.99	122.54			128.59	122.82			139.93
P, S	122.07			243.28	90.31			182.72	131.69			199.86
S, S	10.66			53.92	9.36			48.73	10.68			50.78

TABLE III. Surface tension coefficients (cal mol<sup>-1</sup> Å<sup>-2</sup>) that do not depend on atomic numbers.

SM5.42R/BPW91/MIDI!6D				SM5.42R/BPW91/DZVP				SM5.42R/BPW91/6-31G*			
$\sigma_{CS}^{(\gamma)}$	$\sigma_{CS}^{(\beta^2)}$	$\sigma_{CS}^{(\phi^2)}$	$\sigma_{CS}^{(\psi^2)}$	$\sigma_{CS}^{(\gamma)}$	$\sigma_{CS}^{(\beta^2)}$	$\sigma_{CS}^{(\phi^2)}$	$\sigma_{CS}^{(\psi^2)}$	$\sigma_{CS}^{(\gamma)}$	$\sigma_{CS}^{(\beta^2)}$	$\sigma_{CS}^{(\phi^2)}$	$\sigma_{CS}^{(\psi^2)}$
0.33	3.91	-3.93	-8.41	0.37	-0.15	-4.21	-7.59	0.35	3.51	-4.06	-8.32

Second, and more importantly, electrostatic effects are more important in water than in any organic solvent for which a substantial amount of solvation data is available. Thus, water is a good solvent to use to determine appropriate electrostatic

radii for the solute atoms. The electrostatic radii determined for water will be used without change for organic solvents.

The SM5.42R model has five types of parameters: (i) the intrinsic electrostatic radius for each atomic number used in

TABLE IV. Solvation free energies (kcal/mol) of selected solutes in water computed by SM5.42R/BPW91/MIDI!6D at different solute geometries.

Solute	HF/MIDI!	HF/6-31G*	BPW91/6-31G*	AM1	SM5.4/AM1	Experiment
Ethane	1.12	1.10	1.12	1.11	1.10	1.83
Cyclohexane	0.52	0.45	0.56	0.55	0.55	1.23
1-butene	1.16	1.15	1.20	1.17	1.17	1.38
1-pentyne	0.18	0.17	0.18	0.17	0.15	0.01
Benzene	-1.39	-1.39	-1.47	-1.50	-1.51	-0.87
Ethanol	-5.43	-5.29	-5.50	-5.43	-5.59	-5.01
Dimethyl ether	-2.18	-1.78	-2.18	-2.05	-2.10	-1.92
Tetrahydrofuran	-3.47	-3.42	-3.53	-3.64	-3.67	-3.47
Acetaldehyde	-4.03	-3.98	-4.40	-4.72	-5.03	-3.50
Propanone	-4.30	-4.30	-4.68	-5.25	-5.55	-3.85
Acetic acid	-7.10	-6.90	-7.69	-8.15	-8.57	-6.70
Methyl methanoate	-3.18	-3.08	-3.63	-3.86	-4.11	-2.78
Ethyl ethanoate	-3.23	-3.20	-3.79	-4.16	-4.42	-3.10
<i>m</i> -hydroxybenzaldehyde	-9.60	-9.28	-10.21	-10.28	-10.78	-9.51
Ethylamine	-4.00	-3.87	-3.99	-3.72	-3.81	-4.50
Pyrrolidine	-5.00	-4.97	-4.97	-4.98	-5.11	-5.48
Pyridine	-4.52	-4.47	-4.56	-4.54	-4.58	-4.70
Ethanonitrile	-4.73	-4.75	-5.09	-5.20	-5.24	-3.89
Nitrobenzene	-3.70	-3.94	-4.57	-4.31	-4.70	-4.12
Thiophenol	-2.61	-2.64	-2.65	-2.63	-2.65	-2.55
Dimethyl disulfide	-1.94	-2.00	-2.02	-2.36	-2.43	-1.83
Fluoromethane	-1.15	-1.13	-1.18	-1.08	-1.11	-0.22
Chlorobenzene	-1.55	-1.53	-1.61	-1.54	-1.56	-1.12
2-bromopropene	-0.57	-0.59	-0.59	-0.56	-0.59	-0.86
MUE <sup>a</sup>	0.39 (0.40)	0.37 (0.40)	0.54 (0.44)	0.63 (0.46)	0.74 (0.49)	
MSE <sup>b</sup>	-0.22 (-0.18)	-0.17 (-0.16)	-0.41 (-0.26)	-0.48 (-0.25)	-0.61 (-0.31)	
MUE <sup>c</sup>		0.08 (0.07)	0.20 (0.10)	0.31 (0.17)	0.42 (0.17)	
MSE <sup>d</sup>		0.04 (0.02)	-0.19 (-0.09)	-0.26 (-0.13)	-0.39 (-0.13)	
CH <sub>3</sub> O <sup>-</sup>	-82.7	-88.2	-82.1	-87.7	-87.9	-98
C <sub>6</sub> H <sub>5</sub> O <sup>-</sup>	-63.4	-62.3	-62.3	-62.7	-63.5	-75
CH <sub>3</sub> CH <sub>2</sub> OH <sub>2</sub> <sup>+</sup>	-76.3	-75.7	-75.9	-76.4	-79.1	-81
CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-76.3	-76.4	-76.3	-76.2	-76.4	-73
NO <sub>2</sub> <sup>-</sup>	-75.5	-75.5	-75.3	-75.5	-75.6	-73
(CH <sub>3</sub> ) <sub>2</sub> SH <sup>+</sup>	-62.3	-62.4	-62.2	-62.4	-62.7	-61
MUE <sup>e</sup>	6.4	5.9	6.8	5.7	5.2	
MSE <sup>f</sup>	4.1	3.4	4.5	3.3	2.6	
MUE <sup>g</sup>		1.2	0.4	1.0	1.5	
MSE <sup>h</sup>		-0.7	0.4	-0.7	-1.5	

<sup>a</sup>Mean unsigned error of 24 neutrals against experimental data (kcal/mol) (numbers in parentheses are errors over 18 neutral solutes which do not contain a carbonyl group).

<sup>b</sup>Mean signed error of 24 neutrals against experimental data (kcal/mol) (numbers in parentheses are errors over 18 neutral solutes which do not contain a carbonyl group).

<sup>c</sup>Mean unsigned error of 24 neutrals against solvation free energies computed by SM5.42R/BPW91/MIDI!6D/HF/MIDI! (kcal/mol) (numbers in parentheses are errors over 18 neutral solutes which do not contain a carbonyl group).

<sup>d</sup>Mean signed error of 24 neutrals against solvation free energies computed by SM5.42R/BPW91/MIDI!6D/HF/MIDI! (kcal/mol) (numbers in parentheses are errors over 18 neutral solutes which do not contain a carbonyl group).

<sup>e</sup>Mean unsigned error of six ions against experimental data (kcal/mol).

<sup>f</sup>Mean signed error of six ions against experimental data (kcal/mol).

<sup>g</sup>Mean unsigned error of six ions against solvation free energies computed by SM5.42R/BPW91/MIDI!6D/HF/MIDI! (kcal/mol).

<sup>h</sup>Mean signed error of six ions against solvation free energies computed by SM5.42R/BPW91/MIDI!6D/HF/MIDI! (kcal/mol).



TABLE V. Performance of the SM5.42R model by solvent functional class for nonphosphorus-containing neutral solutes.

Solvent class	Number of			BPW91/MIDI!6D		BPW91/DZVP		BPW91/6-31G*	
	Solvents <sup>a</sup>	Classes <sup>b</sup>	Data <sup>c</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>
Aqueous	1	31	248	-0.02	0.43	0.01	0.41	-0.08	0.44
Alkanes	11	30	475	-0.06	0.33	-0.03	0.32	-0.03	0.31
Cycloalkanes	2	24	106	0.07	0.38	0.14	0.38	0.11	0.37
Arenes	12	16	256	0.19	0.38	0.25	0.38	0.24	0.38
Aliphatic alcohols	12	31	299	-0.05	0.55	0.00	0.57	-0.03	0.52
Aromatic alcohols	2	7	12	0.07	0.55	0.25	0.59	0.15	0.55
Ketones	4	10	35	-0.20	0.49	-0.14	0.43	-0.16	0.44
Esters	2	8	36	0.37	0.60	0.41	0.53	0.37	0.56
Aliphatic ethers	4	19	99	-0.02	0.52	0.07	0.49	0.01	0.49
Aromatic ethers	3	5	15	-0.58	0.59	-0.44	0.53	-0.51	0.54
Amines	2	6	12	0.03	0.36	0.19	0.32	0.10	0.34
Pyridines	3	5	15	-0.31	0.49	-0.17	0.46	-0.24	0.41
Nitriles	2	5	10	-0.76	0.76	-0.66	0.69	-0.67	0.67
Nitrohydrocarbons	4	8	27	-0.30	0.63	-0.36	0.56	-0.28	0.58
Tertamides	2	5	10	-0.14	0.38	-0.03	0.37	-0.06	0.29
Haloaliphatics	12	27	269	-0.12	0.47	-0.10	0.46	-0.09	0.45
Haloaromatics	6	11	106	-0.29	0.40	-0.34	0.44	-0.28	0.39
Misc. acidic	3	5	15	0.02	0.48	0.16	0.49	0.11	0.43
Misc. basic	4	12	39	-0.10	0.42	-0.08	0.41	-0.03	0.39
Total	91	31	2084	-0.04	0.43	0.00	0.43	-0.02	0.41

<sup>a</sup>Number of solvents in this solvent class.<sup>b</sup>Number of solute classes for which data exists in this solvent class.<sup>c</sup>Total number of solute/solvent data involving this solvent class.<sup>d</sup>Mean signed error over data in this solvent class (kcal/mol).<sup>e</sup>Mean unsigned error over data in this solvent class (kcal/mol).

computing the effective atomic radius, (ii) the  $d_{kk'}$  parameters used in the Coulomb integrals in Eq. (3), (iii) the COT parameters in Eqs. (14)–(23), (iv) the atomic radii used in the calculation of the solvent-accessible surface area, and (v) the surface tension coefficient parameters. The parameters of type (iii) and type (iv) were taken to be identical to their values in the SM5.2R model.<sup>12</sup> The type (i) parameters were by default taken to be the same as the SM5.4 model unless the error in affected compounds could be significantly improved by choosing an alternate value. As in earlier SMx models, the type (ii) parameters were adjusted independently for cases where C and H interact ( $d_{CH}$  or  $d_{HC}$ ) and any other

set of interactions ( $d_{kk'}$ ) to minimize the appearance of any positive value for  $\Delta G_{ENP}$ , which is the sum of  $G_p$  and  $\Delta E_{EN}$  in Eq. (1) and becomes  $\Delta G_{EP}$  in R-type models. Type (v) parameters are linear and were found to minimize the error between the solvation free energies predicted with the resulting SM5.42R model and the experimental results for neutral compounds.

In order to determine if modifying a parameter of type (i) could improve the resulting solvation model significantly, we must first define an error function for the model. We selected the error function that was first used for an earlier SMx model<sup>9</sup>

$$U = \frac{\left( \sum_{n=1}^N \left| G_S^0(\text{exp}_n) - G_S^0(\text{calc}_n) \right| + \frac{1}{6} \left( \sum_{i=1}^I \left| G_S^0(\text{exp}_i) - G_S^0(\text{calc}_i) \right| \right) \right)}{N+I}, \quad (26)$$

where  $U$  is the unfitness of a given set of parameters,  $N$  is the number of neutral molecules,  $I$  is the number of ionic compounds,  $G_S^0(\text{exp})$  is the experimental standard-state free energy of solvation, and  $G_S^0(\text{calc})$  is the standard-state free energy of solvation calculated for a given set of nonlinear parameters and reoptimized linear parameters. If varying any one of the intrinsic Coulomb radii improved the unfitness

value of the affected compounds by more than 0.1 kcal/mol, then we adjusted that intrinsic electrostatic radius to minimize the resulting unfitness value. The results of this process will be detailed later in this section. Once all the electrostatic radii were determined, the two  $d_{kk'}$  parameters were determined such that the most positive  $\Delta G_{ENP}$  for any of the molecules in our training set was less than 0.1 kcal/mol.

TABLE VI. Performance of the SM5.42R model by solute functional class for nonphosphorus-containing neutral solutes.

Solute class	Number of			BPW91/MIDI!6D		BPW91/DZVP		BPW91/6-31G*	
	Solutes <sup>a</sup>	Classes <sup>b</sup>	Data <sup>c</sup>	MSE <sup>d</sup>	MSU <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>
Unbranched Alkanes	9	19	84	0.09	0.40	0.21	0.45	0.10	0.39
Branched Alkanes	5	3	12	0.13	0.42	0.08	0.40	0.13	0.43
Cycloalkanes	5	6	18	-0.47	0.52	-0.31	0.35	-0.40	0.46
Alkenes	9	4	27	-0.02	0.28	0.19	0.33	-0.03	0.27
Alkynes	5	3	14	-0.01	0.21	-0.03	0.26	0.01	0.17
Arenes	9	19	134	-0.53	0.57	-0.41	0.50	-0.41	0.48
Alcohols	17	19	385	0.01	0.37	-0.01	0.37	0.02	0.36
Ethers	12	19	93	0.04	0.48	0.11	0.42	0.12	0.45
Aldehydes	7	8	38	-0.08	0.41	0.38	0.55	-0.07	0.40
Ketones	12	18	203	-0.28	0.44	-0.25	0.44	-0.18	0.43
Carboxylic Acids	5	14	124	-0.20	0.45	-0.05	0.41	-0.09	0.42
Esters	14	8	249	0.13	0.33	0.08	0.33	0.11	0.31
Nonhalo bifunctional compounds	5	8	28	0.83	1.07	0.94	1.16	0.78	0.99
Inorganic compounds	2	9	22	-0.01	0.57	0.00	0.60	0.00	0.61
Aliphatic amines	15	10	168	0.08	0.34	0.10	0.36	0.09	0.33
Aromatic amines	11	12	81	-0.01	0.33	0.00	0.30	0.01	0.34
Nitriles	4	6	22	0.00	0.44	0.01	0.48	0.01	0.35
Nitrohydrocarbons	6	8	38	0.01	0.43	0.03	0.55	0.03	0.34
Amides & ureas	4	6	11	0.86	1.24	0.73	1.08	0.74	1.10
Bifunctional HCN and HCNO	6	3	11	-0.43	0.69	-0.43	0.61	-0.37	0.61
Inorganic HCN	2	8	15	-0.42	0.70	-0.35	0.64	-0.36	0.65
Thiols	4	5	14	0.31	0.33	0.24	0.31	0.26	0.29
Sulfides	6	6	23	-0.16	0.59	-0.09	0.61	-0.12	0.57
Disulfides	2	3	5	-0.01	0.25	0.00	0.23	0.00	0.25
Fluorinated hydrocarbons	9	5	19	-0.36	0.49	-0.22	0.35	-0.43	0.48
Chloroalkanes	13	5	35	-0.09	0.34	-0.45	0.49	-0.22	0.33
Chloroalkenes	5	4	16	0.65	0.65	0.65	0.65	0.67	0.67
Chloroarenes	8	6	37	-0.41	0.43	0.11	0.33	-0.32	0.37
Brominated hydrocarbons	14	6	50	-0.27	0.36	-0.20	0.34	-0.48	0.54
Iodinated hydrocarbons	9	6	28	-0.05	0.34	0.01	0.18	-0.03	0.25
Other halo compounds	26	9	80	0.29	0.70	0.30	0.75	0.23	0.70
Total:	260	19	2084	-0.04	0.43	0.00	0.43	-0.02	0.41

<sup>a</sup>Number of solutes in this solute class.<sup>b</sup>Number of solvent classes for which there are data for this solute class.<sup>c</sup>Total number of solute/solvent data involving solutes in this solute class.<sup>d</sup>Mean signed errors over this solute class data (kcal/mol).<sup>e</sup>Mean unsigned errors over this solute class data (kcal/mol).

Our nonphosphorus water training set includes 248 neutral solutes and 43 ions. The parameterization for aqueous solution is broken into three parts. As a first step, we considered only solutes containing at most H, C, N, and O atoms (157 neutral solutes and 24 ions). Using all the nonlinear parameters from the SM5.2R model, we refit the linear surface tension coefficients to minimize the error in the predicted solvation free energies for the 157 neutral solutes. Using this set of parameters we then evaluated the unfitness function in Eq. (26). Since adjusting the intrinsic electrostatic radii for H, C, N, and O never improved the unfitness of the resulting model (with refit linear surface tension coefficients for each choice of trial electrostatic radii) we accepted the previously determined values for the H, C, N, and O intrinsic coulomb radii<sup>13</sup> (see Table I). At this point, we found the values for  $d_{\text{CH}}$ ,  $d_{\text{HC}}$ , and  $d_{\text{kk'}}$  to minimize the occurrence of positive  $\Delta G_{\text{EP}}$ . This yielded  $d_{\text{CH}}=4.2$ ,  $d_{\text{HC}}=4.2$ , and  $d_{\text{kk'}}=3.9$ . With all the nonlinear parameters fixed, the linear surface tension coefficients were refit to minimize the mean unsigned error of the neutrals. Results are listed in Table II.

In the second step for parameterizing the present model

for water, we froze all the parameters determined in the first step and considered the portion of our training set which contains F, S, Cl, Br, and/or I in addition to H, C, O, and N. There are 91 neutral solutes and 9 ions in this portion of the training set. By perturbing the previously published<sup>12</sup> electrostatic radii, refitting the new surface tension coefficients, and recalculating the unfitness function for the molecules in this portion of the parameterization, we determined that no benefit would be changed by altering the intrinsic electrostatic radii of the halogen atoms. Changing the sulfur intrinsic electrostatic radius was found to improve the unfitness of the affected molecules by more than 0.1 kcal/mol, so we adjusted the sulfur radius to minimize the unfitness of the effected molecules. The new sulfur radius is 1.98 Å, and a list of the radii used in this work is in Table I.

As our third step, we performed the parameterization for phosphorus-containing compounds. Our previous experience with phosphorus indicated that it is difficult to obtain good accuracy in the parameterized solvation free energies for phosphorus-containing compounds (errors can be as large as  $\sim 4$  kcal/mol compared to experimental values).<sup>9,12</sup> Since we have made two major improvements over previous SMx sol-

TABLE VII. Calculated and experimental free energies of solvation (kcal/mol) in water for ionic solutes used in the parameterization of the SM5.42R model.

	BPW91/MIDI!6D			BPW91/DZVP $\Delta G_S^0$	BPW91/6-31G* $\Delta G_S^0$	Exp. $\Delta G_S^0$
	$\Delta G_{EP}$	$G_{CDS}$	$\Delta G_S^0$			
HC <sub>2</sub> <sup>-</sup>	-80.3	1.9	-78.4	-78.7	-78.3	-73
CH <sub>3</sub> OH <sub>2</sub> <sup>+</sup>	-83.2	1.1	-82.1	-80.7	-80.7	-87
(CH <sub>3</sub> ) <sub>2</sub> OH <sup>+</sup>	-70.3	0.4	-69.9	-70.0	-69.9	-70
CH <sub>3</sub> CH <sub>2</sub> OH <sub>2</sub> <sup>+</sup>	-77.3	1.0	-76.3	-74.9	-74.9	-81
CH <sub>3</sub> C(OH)CH <sub>3</sub> <sup>+</sup>	-66.8	-0.5	-69.3	-66.6	-66.9	-64
H <sub>3</sub> O <sup>+</sup>	-101.9	5.7	-96.2	-92.8	-93.3	-105
CH <sub>3</sub> O <sup>-</sup>	-82.8	0.1	-82.7	-84.8	-83.7	-98
CH <sub>3</sub> CO <sub>2</sub> <sup>-</sup>	-71.4	0.2	-71.2	-72.9	-71.7	-77
CH <sub>3</sub> COCH <sub>2</sub> <sup>-</sup>	-69.9	0.9	-69.0	-71.4	-69.5	-81
C <sub>6</sub> H <sub>5</sub> O <sup>-</sup>	-63.0	-0.4	-63.4	-66.5	-64.0	-75
C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub> <sup>-</sup>	-58.6	1.2	-57.4	-58.8	-57.2	-59
OH <sup>-</sup>	-102.4	-5.1	-107.5	-110.2	-109.0	-110
HO <sub>2</sub> <sup>-</sup>	-89.9	-6.8	-96.7	-98.8	-98.6	-101
O <sub>2</sub> <sup>-</sup>	-86.8	-6.0	-92.8	-93.9	-94.0	-87
CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-74.6	-1.7	-76.3	-76.3	-59.2	-73
HCONH <sub>3</sub> <sup>+</sup>	-73.4	-4.2	-77.6	-76.4	-76.8	-78
CH <sub>3</sub> CNH <sup>+</sup>	-67.8	1.3	-66.5	-66.1	-65.7	-69
CH <sub>3</sub> C(OH)NH <sub>2</sub> <sup>+</sup>	-66.7	-4.6	-71.3	-69.9	-70.3	-70
(CH <sub>3</sub> ) <sub>2</sub> NH <sub>2</sub> <sup>+</sup>	-67.2	0.1	-67.1	-67.1	-66.9	-66
(CH <sub>3</sub> ) <sub>3</sub> NH <sup>+</sup>	-61.6	2.3	-59.3	-59.2	-59.2	-59
Imidazole·H <sup>+</sup>	-61.4	-1.0	-62.4	-61.1	-61.6	-64
Pyridine·H <sup>+</sup>	-58.5	-0.2	-58.7	-57.9	-58.7	-58
C <sub>6</sub> H <sub>5</sub> NH <sub>3</sub> <sup>+</sup>	-66.0	-1.9	-67.9	-68.2	-68.1	-68
NH <sub>4</sub> <sup>+</sup>	-84.3	-3.0	-87.3	-87.4	-87.0	-81
CN <sup>-</sup>	-77.8	3.5	-74.3	-74.5	-74.3	-75
CH <sub>2</sub> CN <sup>-</sup>	-67.4	2.0	-65.4	-66.7	-65.6	-75
NH <sub>2</sub> <sup>-</sup>	-86.0	-2.3	-88.3	-89.4	-89.1	-95
NO <sub>2</sub> <sup>-</sup>	-74.9	-0.6	-75.5	-75.0	-75.6	-73
NO <sub>3</sub> <sup>-</sup>	-69.1	0.1	-69.0	-67.6	-69.1	-66
N <sub>3</sub> <sup>-</sup>	-66.5	-2.8	-69.3	-71.0	-71.4	-74
CH <sub>3</sub> SH <sub>2</sub> <sup>+</sup>	-73.0	2.0	-71.0	-68.7	-69.9	-74
(CH <sub>3</sub> ) <sub>2</sub> SH <sup>+</sup>	-64.6	2.3	-62.3	-61.8	-62.0	-61
HS <sup>-</sup>	-82.4	-1.8	-84.2	-85.2	-84.4	-76
CH <sub>3</sub> S <sup>-</sup>	-75.4	-1.4	-76.8	-79.1	-79.6	-76
CH <sub>3</sub> CH <sub>2</sub> S <sup>-</sup>	-71.5	-1.5	-73.0	-75.0	-75.8	-74
<i>n</i> -C <sub>3</sub> H <sub>7</sub> S <sup>-</sup>	-70.0	-1.4	-71.4	-73.1	-74.4	-76
C <sub>6</sub> H <sub>5</sub> S <sup>-</sup>	-61.3	-2.2	-63.5	-63.2	-64.8	-65
F <sup>-</sup>	-109.3	1.3	-108.0	-108.3	-108.2	-107
CHF <sub>2</sub> CO <sub>2</sub> <sup>-</sup>	-65.7	1.4	-64.3	-64.9	-64.3	-70
Cl <sup>-</sup>	-77.0	-0.2	-77.2	-77.2	-77.0	-78
CHCl <sub>2</sub> CO <sub>2</sub> <sup>-</sup>	-58.3	-0.3	-58.6	-59.3	-58.0	-66
Br <sup>-</sup>	-71.0	-0.9	-71.9	-71.9	-72.6	-72
I <sup>-</sup>	-61.6	-1.3	-62.9	-62.8	-62.8	-63
Mean unsigned error:			3.9	3.6	3.9	
Mean signed error:			1.8	1.5	1.0	

vation models by incorporating the DFT Hamiltonian and the CM2 charge scheme<sup>26</sup> in SM5.42R, it is interesting to see how the present model performs for solutes that contain phosphorus. Based on 20 aqueous experimental solvation free energies (14 for neutral solutes, 6 for ions),<sup>12</sup> we minimized the unfitness value for the phosphorus compounds and found the optimal intrinsic electrostatic radius for P to be 2.27 Å. The three P surface tension coefficients were then reoptimized using the new electrostatic radius. This completed the parameterization for water and the resulting data in Table II.

### III.E. Parameterization for organic solvents

We used a total of 1836 experimental solvation free energies of 260 neutral solutes composed of H, C, N, O, F, S, Cl, Br, and I atoms in 90 organic solvents to parameterize nonphosphorus solvation energies in organic solvents for phosphorus parameterization. We have 37 free energy data points for 15 phosphorus-containing solutes in 9 organic solvents. Since the atomic surface tension coefficients depend on several solvent descriptors, the parameters we were trying to optimize are those associated with the descriptors in Eqs. (24) and (25). The values of all the electrostatic radii were

TABLE VIII. Solvation free energies (kcal/mol) for selected neutral solutes in selected solvents.

Solute	Method	Solvent						
		Hexadecane	Cyclohexane	Benzene	1-octanol	Diethylether	Chloroform	Water
<i>n</i> -octane	SM5.42R/BPW91/MIDI!6D	-4.62	-4.81	-5.61	-3.79	-5.04	-5.42	2.39
	SM5.42R/BPW91/DZVP	-4.57	-4.79	-5.55	-3.57	-4.98	-5.25	2.46
	SM5.42R/BPW91/6-31G*	-4.56	-4.77	-5.59	-3.84	-5.09	-5.37	2.27
	SM5.4/AM1	-4.61	-4.76	-5.55	-4.40	-4.99	-5.07	2.15
	Experimental	-5.02	-5.63	-5.35	-4.18	-5.62	-5.25	-2.89
Benzene	SM5.42R/BPW91/MIDI!6D	-4.54	-4.62	-5.21	-4.55	-5.21	-5.42	-1.39
	SM5.42R/BPW91/DZVP	-4.24	-4.33	-4.86	-4.25	-4.96	-5.21	-1.04
	SM5.42R/BPW91/6-31G*	-4.48	-4.57	-5.18	-4.50	-5.16	-5.30	-1.30
	SM5.4/AM1	-3.86	-3.95	-4.53	-3.87	-4.67	-4.69	-0.97
	Experiment	-3.80	-4.19	-4.55	-3.72	-4.21	-4.64	-0.87
Toluene	SM5.42R/BPW91/MIDI!6D	-4.97	-5.07	-5.75	-4.99	-5.76	-6.00	-1.24
	SM5.42R/BPW91/DZVP	-4.80	-4.91	-5.53	-4.89	-5.69	-5.97	-1.12
	SM5.42R/BPW91/6-31G*	-4.88	-4.99	-5.68	-4.88	-5.67	-5.82	-1.09
	SM5.4/AM1	-4.52	-4.61	-5.28	-4.66	-5.38	-5.38	-0.89
	Experiment	-4.54	-4.90	-5.32	-4.55	-5.23	-5.48	-0.89
1-butanol	SM5.42R/BPW91/MIDI!6D	-3.46	-3.55	-4.47	-6.02	-5.36	-5.13	-4.65
	SM5.42R/BPW91/DZVP	-3.39	-3.50	-4.35	-5.80	-5.30	-5.10	-4.62
	SM5.42R/BPW91/6-31G*	-3.37	-3.47	-4.37	-5.93	-5.36	-5.13	-4.69
	SM5.4/AM1	-3.39	-3.47	-4.31	-5.56	-5.11	-4.95	-4.39
	Experiment	-3.55	-3.52	-4.45	-5.71	-5.69	-5.28	-4.72
Phenol	SM5.42R/BPW91/MIDI!6D	-5.16	-5.23	-6.23	-7.83	-7.22	-7.04	-6.32
	SM5.42R/BPW91/DZVP	-5.09	-5.16	-6.09	-8.00	-7.40	-7.36	-6.62
	SM5.42R/BPW91/6-31G*	-5.16	-5.23	-6.23	-7.91	-7.35	-7.18	-6.46
	SM5.4/AM1	-5.47	-5.52	-6.45	-7.91	-7.71	-7.21	-6.55
	Experiment	-5.14	-5.57	-7.12	-8.69	-8.75	-7.14	-6.62
1,4-dioxane	SM5.42R/BPW91/MIDI!6D	-3.93	-4.01	-4.66	-5.29	-5.36	-5.62	-5.67
	SM5.42R/BPW91/DZVP	-4.00	-4.09	-4.73	-5.15	-5.22	-5.44	-5.38
	SM5.42R/BPW91/6-31G*	-3.96	-4.04	-4.72	-5.14	-5.30	-5.45	-5.64
	SM5.4/AM1	-3.65	-3.73	-4.47	-5.49	-5.13	-6.31	-6.03
	Experiment	-3.82	-4.17	-5.21	-4.89	-4.67	-6.21	-5.05
Butanone	SM5.42R/BPW91/MIDI!6D	-3.59	-3.66	-4.26	-5.19	-5.07	-5.56	-3.91
	SM5.42R/BPW91/DZVP	-3.49	-3.57	-4.15	-5.21	-4.97	-5.48	-3.98
	SM5.42R/BPW91/6-31G*	-3.50	-3.57	-4.17	-5.08	-4.96	-5.46	-3.84
	SM5.4/AM1	-3.31	-3.38	-4.07	-4.96	-4.83	-5.00	-3.62
	Experiment	-3.12	-3.48	-4.46	-3.78	-4.09	-5.43	-3.64
Propanoic acid	SM5.42R/BPW91/MIDI!6D	-3.96	-4.02	-4.92	-6.82	-6.26	-5.97	-6.50
	SM5.42R/BPW91/DZVP	-3.65	-3.72	-4.60	-6.70	-6.16	-5.78	-6.35
	SM5.42R/BPW91/6-31G*	-3.72	-3.78	-4.69	-6.74	-6.21	-5.86	-6.36
	SM5.4/AM1	-2.94	-3.00	-3.98	-6.22	-6.02	-5.22	-5.85
	Experiment	-3.12	-3.78	-4.75	-6.86	-6.75	-5.37	-6.47
Butylamine	SM5.42R/BPW91/MIDI!6D	-3.45	-3.56	-4.28	-4.93	-4.47	-4.77	-3.61
	SM5.42R/BPW91/DZVP	-3.45	-3.58	-4.26	-4.86	-4.41	-4.70	-3.61
	SM5.42R/BPW91/6-31G*	-3.45	-3.57	-4.29	-4.94	-4.49	-4.75	-3.66
	SM5.4/AM1	-3.95	-4.04	-4.71	-5.39	-4.80	-5.52	-4.26
	Experiment	-3.57	-3.72	-4.33	-5.35	-4.44	-5.35	-4.29
pyridine	SM5.42R/BPW91/MIDI!6D	-4.60	-4.66	-5.15	-5.62	-5.29	-6.20	-4.52
	SM5.42R/BPW91/DZVP	-4.41	-4.48	-4.95	-5.48	-5.23	-6.06	-4.31
	SM5.42R/BPW91/6-31G*	-4.63	-4.70	-5.19	-5.58	-5.26	-6.15	-4.55
	SM5.4/AM1	-4.54	-4.59	-5.01	-6.24	-5.40	-6.56	-5.94
	Experiment	-4.10	-4.30	-5.28	-5.34	-4.81	-6.45	-4.70

TABLE VIII. *Continued*

Solute	Method	Solvent						
		Hexadecane	Cyclohexane	Benzene	1-octanol	Diethylether	Chloroform	Water
Aniline	SM5.42R/BPW91/MIDI!6D	-5.50	-5.58	-6.44	-7.14	-7.04	-7.15	-5.57
	SM5.42R/BPW91/DZVP	-5.45	-5.53	-6.35	-7.22	-7.08	-7.28	-5.62
	SM5.42R/BPW91/6-31G*	-5.48	-5.57	-6.45	-7.15	-7.03	-7.10	-5.55
	SM5.4/AM1	-5.71	-5.78	-6.61	-7.16	-7.19	-7.15	-5.14
	Experiment	-5.44	-5.52	-6.88	-6.71	-6.51	-7.34	-5.49
Nitrobenzene	SM5.42R/BPW91/MIDI!6D	-5.85	-5.94	-6.64	-5.93	-6.85	-7.26	-3.70
	SM5.42R/BPW91/DZVP	-5.62	-5.71	-6.33	-5.59	-6.79	-7.22	-3.28
	SM5.42R/BPW91/6-31G*	-5.89	-5.98	-6.67	-6.06	-6.91	-7.38	-3.86
	SM5.4/AM1	-5.88	-5.92	-6.43	-5.51	-7.10	-7.75	-3.07
	Experiment	-6.22	-6.62	-7.60	-6.63	-6.85	-7.78	-4.12
Thiophene	SM5.42R/BPW91/MIDI!6D	-3.93	-3.99	-4.31	-3.91	-3.97	-4.91	-2.09
	SM5.42R/BPW91/DZVP	-3.82	-4.00	-4.34	-3.81	-4.05	-4.87	-1.97
	SM5.42R/BPW91/6-31G*	-3.84	-3.92	-4.26	-3.80	-3.80	-4.78	-1.83
	SM5.4/AM1	-4.91	-4.99	-5.48	-4.89	-5.14	-5.28	-2.90
	Experiment	-4.01	n.a.	n.a.	-3.89	n.a.	-5.83	-1.42
Chlorobenzene	SM5.42R/BPW91/MIDI!6D	-5.21	-5.30	-5.97	-5.25	-5.90	-6.19	-1.55
	SM5.42R/BPW91/DZVP	-4.86	-4.97	-6.03	-4.82	-5.56	-5.85	-1.04
	SM5.42R/BPW91/6-31G*	-5.14	-5.24	-5.93	-5.19	-5.85	-6.07	-1.43
	SM5.4/AM1	-4.81	-4.90	-5.57	-4.86	-5.55	-5.43	-1.16
	Experiment	-4.99	-5.10	n.a.	-5.00	-5.42	-5.45	-1.12

taken from those in the water parameterization, and the surface tension coefficients were optimized by a least squares method on the error of the neutrals. The solutes that contain only H, C, N, and O atoms were used first in the parameterization to obtain the surface tension coefficients for these four atomic types. The data set for H, C, N, and O comprises 1620 experimental solvation free energies of 164 solutes in all 90 solvents. The parameters for H, C, N, and O atoms were then fixed and used in the parameterization for solutes that also contain sulfur and halogen atoms (216 free energy data points of 96 solutes in 34 organic solvents). All these parameters were frozen in the parameterization for phosphorus. Results are summarized in Tables II and III.

## IV. RESULTS AND DISCUSSION

### IV.A. Comparison of solvation energies based on different solute geometries

The essence of R-type solvation models is that they should be able to predict solvation free energies accurately based on any reasonably accurate solute geometries. It is therefore interesting to see how the computed solvation free energies change when different solute geometries are used in the present model. In Table IV, we present the calculated solvation free energies of a set of selected solutes (24 neutrals plus 6 ions) in water using the SM5.42R/BPW91/MIDI!6D parameterization. Five different sets of solute ge-

TABLE IX. Performance of the SM5.42R model on phosphorus-containing neutral solutes by solvent functional class.

Solvent class	Number of			BPW91/MIDI!6D		BPW91/DZVP		BPW91/6-31G*	
	Solvents <sup>a</sup>	Classes <sup>b</sup>	Data <sup>c</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>
Aqueous	1	2	14	0.26	0.83	0.03	1.02	0.02	0.89
Alkanes	3	1	7	-0.33	1.15	-0.23	1.13	-0.44	1.28
Cycloalkanes	1	1	3	0.02	0.85	0.16	0.85	-0.11	0.83
Arenes	1	2	6	0.10	1.30	0.18	1.26	-0.13	1.31
Aliphatic alcohols	1	2	12	0.41	1.30	0.13	1.22	0.31	1.32
Haloaliphatics	3	2	9	-0.38	1.35	-0.05	1.09	0.02	1.32
Total:	10	2	51	0.07	1.13	0.03	1.11	0.00	1.17

<sup>a</sup>Number of solvents in this solvent class in which phosphorus-containing solutes are represented in training set.

<sup>b</sup>Number of solute classes represented in this solvent class. The phosphorus solute classes are listed in Table X.

<sup>c</sup>Total number of data for phosphorus-containing solutes in this solvent class.

<sup>d</sup>Mean signed errors over data in this solvent class (kcal/mol).

<sup>e</sup>Mean unsigned errors over data in this solvent class (kcal/mol).

TABLE X. Performance of the SM5.42R model on phosphorus-containing neutral solutes by solute class.

Solute class	Number of			BPW91/MIDI!6D		BPW91/DZVP		BPW91/6-31G*	
	Solutes <sup>a</sup>	Classes <sup>b</sup>	Data <sup>c</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>
P, H, C, and O compounds	5	6	28	−0.05	0.97	−0.02	0.92	−0.03	1.03
Other P compounds	10	4	23	0.21	1.33	0.10	1.35	0.04	1.33
Total:	15	6	51	0.07	1.13	0.03	1.11	0.00	1.17

<sup>a</sup>Number of solutes in this solute class.<sup>b</sup>Number of solvent classes in which this solute class is represented. The solvent classes are listed in Table V.<sup>c</sup>Total number of data for this solute class.<sup>d</sup>Mean signed errors over data in this solute class (kcal/mol).<sup>e</sup>Mean unsigned errors over data in this solute class (kcal/mol).

ometries were used in these calculations. They were optimized at the levels of HF/MIDI! (those geometries were used in the parameterization), HF/6-31G\*, BPW91/6-31G\*, Austin Model 1 (AM1),<sup>63</sup> and SM5.4/AM1.<sup>6–8,10</sup> The mean signed and unsigned errors in solvation free energies for these solutes as compared to experimental data and also as compared to those obtained using the HF/MIDI! geometries were calculated and listed in the table.

We observe the following: First, the results based on HF geometries are remarkably insensitive to the basis set used in the geometry optimization step. Second, the computed solvation free energies based on DFT geometries are very close to those obtained based on HF geometries. For neutrals, the differences are about 0.2 kcal/mol while for ions they are around 1 kcal/mol. These differences are only large for carbonyl compounds, for which the DFT method gives larger bond lengths and hence more polar C=O bonds. Third, there is little change ( $\sim 0.1$  kcal/mol increase in the mean deviation from experiment) in the computed solvation free energies when even much-lower-quality gas-phase AM1 solute geometries are used instead of gas-phase DFT geometries. Fourth, when gas-phase AM1 geometries are replaced by the solution-phase geometries optimized using the SM5.4/AM1 model, the mean deviation from either experiment or the values obtained with the geometries used for parameterization is only around 0.1 kcal/mol for the neutrals and 0.5 kcal/mol for the ions. It is clear from this table that any reasonably accurate solute geometries can be used with the present R models.

#### IV.B. Performance for solutes without phosphorus

The performance for nonphosphorus-containing compounds will be considered first. For 260 neutral solutes in 91 solvents (a total of 2084 data points), the mean unsigned deviations from experimental solvation free energies are 0.43 kcal/mol for both SM5.42R/BPW91/MIDI!6D and SM5.42R/BPW91/DZVP, and 0.41 kcal/mol for SM5.42R/BPW91/6-31G\*. For the 43 ionic solutes in our test set, both SM5.42R/BPW91/MIDI!6D and SM5.42R/BPW91/6-31G\* yield mean unsigned errors of 3.9 kcal/mol, while SM5.42R/BPW91/DZVP gives 3.6 kcal/mol. These errors are similar to those found in previous SMx aqueous and organic solvation models and are also comparable with experimental uncertainties.

*Performance by solvent.* Table V presents the mean errors for 19 solvent functional group classes for the SM5.42R DFT model. For most of the solvent classes, the mean unsigned errors range from 0.3 kcal/mol to 0.6 kcal/mol. The largest errors are for the 10 nitrile compounds where the magnitude of the mean signed and unsigned errors are about 0.7 kcal/mol. Clearly we have systematic errors for nitriles. This is also the case for aromatic ethers. The three DFT solvation parameterizations perform particularly well for alkanes, arenes, amines, and tertiary amides with all mean unsigned errors around 0.35 kcal/mol. Comparing the errors of the three parameterizations for all the solvent classes shows that they are very similar. For example, for the 248 experimental solvation free energies in the test set for water, the mean signed and unsigned errors are  $-0.02$  kcal/mol and  $0.42$  kcal/mol, respectively, for SM5.42R/BPW91/MIDI!6D,  $-0.01$  kcal/mol and  $0.41$  kcal/mol, for SM5.42R/BPW91/DZVP, and  $-0.01$  kcal/mol and  $0.44$  kcal/mol, for SM5.42R/BPW91/6-31G\*. This similarity between the three parameterizations can be found in almost all solvent classes.

*Performance by solutes.* Table VI shows the performance of the SM5.42R DFT solvation model for 31 solute functional group classes. In most cases, the mean unsigned errors in the three parameterizations are between 0.3 kcal/mol and 0.7 kcal/mol. The three parameterizations all have large errors (around 1.1 kcal/mol) for nonhalo bifunctional solutes and amides as well as ureas. In general, for a given solute class, the three parameterizations produce similar signed and unsigned errors. We also notice that for solute classes containing more than 80 data points, the mean unsigned errors are always smaller than 0.5 kcal/mol.

*Performance for ions.* In Table VII, we present the calculated values for  $\Delta G_{EP}$ ,  $G_{CDS}$ , and  $\Delta G_S^0$  for the 43 ions. Compared to experimental solvation free energies for these ions, both SM5.42R/BPW91/MIDI!6D and SM5.42R/BPW91/6-31G\* yield mean unsigned errors of 3.9 kcal/mol and SM5.42R/BPW91/DZVP gives 3.6 kcal/mol. For ions, the solvation free energies are dominated by the solute-solvent electrostatic interactions. We find relatively large errors (over 10 kcal/mol) for a few ions, including  $\text{CH}_3\text{O}^-$ ,  $\text{CH}_3\text{COCH}_2^-$ , and  $\text{C}_6\text{H}_5\text{O}^-$ . In general, we observe that the model has relatively larger errors for anions than for cations. As in the case of neutral solutes, the three DFT parameterizations give very similar solvation free energies for the same ion.

TABLE XI. Calculated and experimental free energies of solvation (kcal/mol) in water for phosphorus-containing ionic solutes.

	BPW91/MIDI!6D			BPW91/DZVP $\Delta G_S^0$	BPW91/6-31G* $\Delta G_S^0$	Exp. $\Delta G_S^0$
	$\Delta G_{EP}$	$G_{CDS}$	$\Delta G_S^0$			
$\text{PH}_4^+$	-71.9	2.2	-69.7	-69.6	-69.4	-73
$\text{CH}_3\text{PH}_3^+$	-65.3	2.9	-62.4	-64.6	-64.5	-63
$(\text{CH}_3)_2\text{PH}_2^+$	-60.5	3.6	-56.9	-60.0	-60.1	-57
$(\text{CH}_3)_3\text{PH}^+$	-57.0	4.3	-52.7	-55.9	-56.2	-53
$\text{PH}_2$	-72.2	-0.5	-72.7	-72.6	-72.3	-67
$\text{H}_2\text{PO}_4^-$	-67.8	-5.8	-73.6	-75.0	-73.5	-68
Mean unsigned error			2.6	3.9	3.7	
Mean signed error			-1.2	-2.8	-2.5	

*Performance for selected solutes and solvents.* To further demonstrate the performance of the DFT solvation model, in Table VIII we listed the computed solvation free energies for 14 selected solutes in 8 representative solvents. These results are compared with those obtained from the SM5.4/AM1 solvation model<sup>6,7</sup> and experimental values. The solvation free energies calculated by the SM5.42R DFT model generally agree well with experimental values and those from the SM5.4/AM1 model, which is based on the CM1 charge scheme. This good agreement of different parameterizations is very encouraging.

#### IV.C. Performance for phosphorus-containing compounds

In the previously developed SM2.2PD,<sup>9</sup> SM5.2PD,<sup>9</sup> and SM5.2R<sup>12</sup> solvation models, we found that the calculated solvation free energies of phosphorus-containing compounds

have large errors. Both SM2.2PD/AM1 and SM5.2PD/AM1 give mean unsigned errors of  $\sim 4$  kcal/mol for 14 neutral phosphorus solutes in water.<sup>10</sup> The mean unsigned errors are 1.0–2.0 kcal/mol in SM5.2R when semiempirical geometries are used for phosphorus solutes; even with HF/MIDI! geometries, SM5.2R still has mean unsigned errors as large of 1.1–1.4 kcal/mol.<sup>12</sup> The performance of the present model for 15 neutral phosphorus compounds is shown in Table IX by solvent functional class and in Table X by solute class. We observe that SM5.42R/BPW91/MIDI!6D gives a total mean unsigned error of 1.13 kcal/mol, SM5.42R/BPW91/DZVP has an error of 1.11 kcal/mol, and SM5.42R/BPW91/6-31G\* yields 1.17 kcal/mol. Although these errors are still relatively large compared with those for most of the solute functional classes in Table VII, the present model, along with the MNDO/*d* solvation of Ref. 12, is the best solvation model that has been created for phosphorus-

TABLE XII. Performance of the SM5.42R model for neutral solutes by solvent functional class.

Solvent class	Number of			BPW91/MIDI!6D		BPW91/DZVP		BPW91/6-31G*	
	Solvents <sup>a</sup>	Classes <sup>b</sup>	Data <sup>c</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>
Aqueous	1	32	262	-0.01	0.45	-0.01	0.45	-0.09	0.47
Alkanes	11	31	482	-0.07	0.34	-0.03	0.33	-0.03	0.32
Cycloalkanes	2	25	109	0.07	0.39	0.14	0.39	0.10	0.39
Arenes	12	17	262	0.19	0.40	0.25	0.40	0.23	0.40
Aliphatic alcohols	12	32	311	-0.03	0.58	0.01	0.59	-0.02	0.55
Aromatic alcohols	2	7	12	0.07	0.55	0.25	0.59	0.15	0.55
Ketones	4	10	35	-0.20	0.49	-0.14	0.43	-0.16	0.44
Esters	2	8	36	0.37	0.60	0.41	0.53	0.37	0.56
Aliphatic ethers	4	19	99	-0.02	0.52	0.07	0.49	0.01	0.49
Aromatic ethers	3	5	15	-0.58	0.59	-0.44	0.53	-0.51	0.54
Amines	2	6	12	0.03	0.36	0.19	0.32	0.10	0.34
Pyridines	3	5	15	-0.31	0.49	-0.17	0.46	-0.24	0.41
Nitriles	2	5	10	-0.76	0.76	-0.66	0.69	-0.67	0.67
Nitrohydrocarbons	4	8	27	-0.30	0.63	-0.36	0.56	-0.28	0.58
Tertamides	2	5	10	-0.14	0.38	-0.03	0.37	-0.06	0.29
Haloaliphatics	12	28	278	-0.13	0.50	-0.09	0.49	-0.09	0.47
Haloaromatics	6	11	106	-0.29	0.40	-0.34	0.44	-0.28	0.39
Misc. acidic	3	5	15	0.02	0.48	0.16	0.49	0.11	0.43
Misc. basic	4	12	39	-0.10	0.42	-0.08	0.41	-0.03	0.39
Total	91	32	2135	-0.04	0.45	0.00	0.44	-0.02	0.43

<sup>a</sup>Number of solvents in this solvent class.

<sup>b</sup>Number of solute classes for which data exists in this solvent class.

<sup>c</sup>Total number of solute/solvent data involving this solvent class.

<sup>d</sup>Mean signed errors over data in this solvent class (kcal/mol).

<sup>e</sup>Mean unsigned errors over data in this solvent class (kcal/mol).

TABLE XIII. Performance of the SM5.42R model for neutral solutes by solute functional class.

Solvent class	Number of			BPW91/MIDI!6D		BPW91/DZVP		BPW91/6-31G*	
	Solutes <sup>a</sup>	Classes <sup>b</sup>	Data <sup>c</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>
Unbranched Alkanes	9	19	84	0.09	0.40	0.21	0.45	0.10	0.39
Branched Alkanes	5	3	12	0.13	0.42	0.08	0.40	0.13	0.43
Cycloalkanes	5	6	18	-0.47	0.52	-0.31	0.35	-0.40	0.46
Alkenes	9	4	27	-0.02	0.28	0.19	0.35	-0.03	0.27
Alkynes	5	3	14	-0.01	0.21	-0.03	0.26	0.01	0.17
Arenes	9	19	134	-0.53	0.57	-0.41	0.50	-0.41	0.48
Alcohols	17	19	385	0.01	0.37	-0.01	0.37	0.02	0.36
Ethers	12	19	93	0.04	0.48	0.11	0.42	0.12	0.45
Aldehydes	7	8	38	-0.08	0.41	0.38	0.55	-0.07	0.40
Ketones	12	18	203	-0.28	0.44	-0.25	0.44	-0.18	0.43
Carboxylic Acids	5	14	124	-0.20	0.45	-0.05	0.41	-0.09	0.42
Esters	14	8	249	0.12	0.33	0.07	0.34	0.10	0.31
Nonhalo bifunctional compounds	5	8	28	0.83	1.07	0.94	1.16	0.78	0.99
Inorganic Compounds	2	9	22	-0.01	0.57	0.00	0.60	0.00	0.61
Aliphatic Amines	15	10	168	0.08	0.34	0.10	0.36	0.09	0.33
Aromatic Amines	11	12	81	-0.01	0.33	-0.06	0.32	0.01	0.34
Nitriles	4	6	22	0.00	0.44	0.00	0.48	0.01	0.35
Nitrohydrocarbons	6	8	38	0.01	0.43	0.03	0.55	0.03	0.34
Amides & Ureas	4	6	11	0.86	1.24	0.73	1.08	0.74	1.10
Bifunctional HCN and HCNO	6	3	11	-0.43	0.69	-0.43	0.61	-0.37	0.61
Inorganic HCN	2	8	15	-0.42	0.70	-0.03	0.64	-0.36	0.65
Phosphorus containing	15	6	51	0.07	1.13	0.25	1.11	0.00	1.17
Thiols	4	5	14	0.31	0.33	0.25	0.31	0.26	0.29
Sulfides	6	6	23	-0.16	0.59	-0.09	0.61	-0.12	0.57
Disulfides	2	3	5	-0.01	0.25	0.00	0.23	0.00	0.25
Fluorinated Hydrocarbons	9	5	19	-0.37	0.49	-0.22	0.35	-0.43	0.48
Chloroalkanes	13	5	35	-0.09	0.34	-0.45	0.49	-0.22	0.33
Chloroalkenes	5	4	16	0.65	0.65	0.65	0.65	0.67	0.67
Chloroarenes	8	6	37	-0.41	0.43	0.11	0.33	-0.32	0.37
Brominated hydrocarbons	14	6	50	-0.27	0.36	-0.20	0.34	-0.48	0.54
Iodinated hydrocarbons	9	6	28	-0.05	0.34	-0.01	0.18	-0.03	0.25
Other halo compounds	26	9	80	0.29	0.70	0.30	0.75	0.23	0.70
Total:	275	19	2135	-0.04	0.45	0.00	0.44	-0.02	0.43

<sup>a</sup>Number of solutes in this solute class.<sup>b</sup>Number of solvent classes for which there are data for this solute class.<sup>c</sup>Total number of solute/solvent data involving solutes in this solute class.<sup>d</sup>Mean signed errors over this solute class data (kcal/mol).<sup>e</sup>Mean unsigned errors over this solute class data (kcal/mol).

containing compounds. The success of the models including  $d$  functions vindicates the considerable effort we put into developing solvation models with  $d$  orbitals. In Table XI are listed the calculated solvation free energies for the six phosphorus ions used in the parameterization; mean unsigned errors are less than 4 kcal/mol, which is excellent accuracy for ions.

#### IV.D. OVERALL PERFORMANCE

The errors in solvation free energies for all 275 neutral solutes in the SM5.42R model are shown by solvent class in Table XII, and by solute class in Table XIII. SM5.42R/BPW91/MIDI!6D, SM5.42R/BPW91/DZVP, and SM5.42R/BPW91/6-31G\* yield mean unsigned errors of 0.45 kcal/mol, 0.44 kcal/mol, and 0.43 kcal/mol, respectively.

#### IV.E. Surface tension coefficients

The fact that the three parameterizations of the present model perform so similarly has led us to study the surface

tension coefficients optimized for them. As we look at Tables II and III, we find that in most cases the surface tension coefficients obtained for the three parameterizations are not significantly different. This is particularly evident for the surface tension coefficients for single atom types (i.e., the first nine rows in Table II). However, in some cases the values are quite different. For example, the optimized surface tension coefficient for H bonded to O in the aqueous SM5.42R/BPW91/MIDI!6D parameterization is almost twice as large as the one in the SM5.42R/BPW91/DZVP. The only difference in our three DFT parameterizations is the basis set employed. Although the CM2 charge model tends to give similar atomic charges for different basis sets, some bond-specific surface tension coefficients are affected in a subtle way.

#### IV.F. Examples

In Fig. 1, we show the CM2 partial atomic charges in the gas phase and water, and atomic GB polarization energies



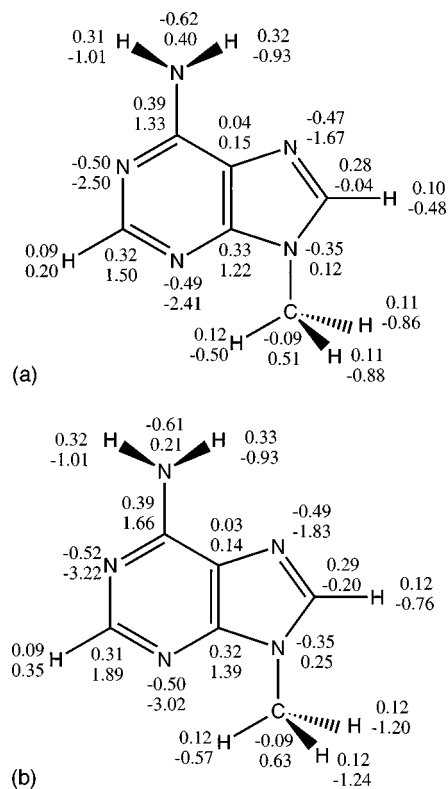


FIG. 1. CM2 partial atomic charges (a.u.) in the gas phase and water, and atomic GB polarization energies (kcal/mol) for 9-methyladenine (HF/MIDI! geometry) in water computed with these charges. There are two numbers for each atom. The top one is the CM2 charge, and the bottom one is the atomic GB polarization energy. Part (a) shows the gas-phase charges and associated  $G_p^k$  in water, and part (b) shows the SCRF charges and  $G_p^k$  in water.

computed with these charges by SM5.42R/BPW91/MIDI!D for 9-methyladenine. Based on Eq. (2), the atomic GB polarization energy for atom  $k$  is defined as:

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

Figure 2 shows the same types of data for 9-methyladenine in 1-octanol. The two figures show that the CM2 charges of 9-methyladenine do not change much going from gas phase to either water or octanol (all changes are less than 0.02 a.u.). For most atoms, the magnitude of the GB polarization energies increases, and the change can be larger than 1 kcal/mol. The solvation free energies calculated at the SM5.42R/BPW91/MIDI!6D level are  $-13.65$  kcal/mol in water and  $-13.54$  kcal/mol in octanol, respectively. These agree very well with the experimental values<sup>55</sup> of  $-13.60$  kcal/mol and  $-13.56$  kcal/mol.

## V. CONCLUDING REMARKS

We have presented a DFT-based SM5-type solvation model called SM5.42R, and we have parameterized it for three basis sets. We have shown that the three parameterizations yield accurate solvation free energies in both aqueous and organic solvents. Our DFT solvation model has the following advantages: (1) DFT provides a more reliable way to describe the solute electronic structure than semiempirical methods. In addition, with gradient-corrected functionals,

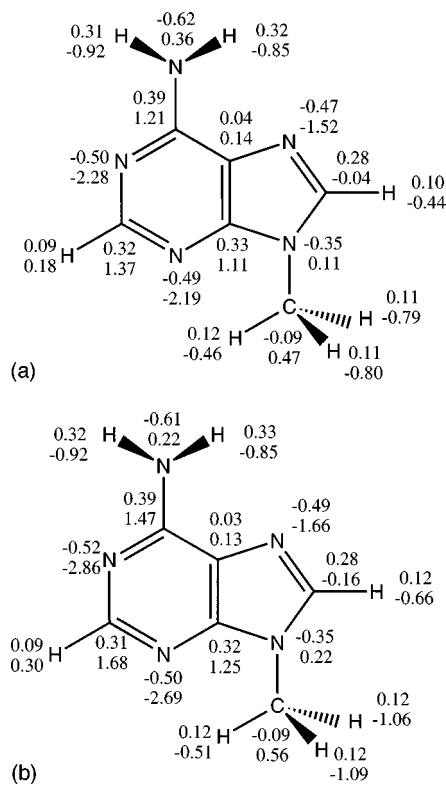


FIG. 2. CM2 partial atomic charges (a.u.) in the gas phase and 1-octanol, and atomic GB polarization energies (kcal/mol) for 9-methyladenine (HF/MIDI! geometry) in 1-octanol computed with these charges. There are two numbers for each atom. The top one is the CM2 charge, and the bottom one is the atomic GB polarization energy. Part (a) shows the gas-phase charges and associated  $G_p^k$  in 1-octanol, and part (b) shows the SCRF charges and  $G_p^k$  in 1-octanol.

reasonably accurate account is taken of electron correlation. (2) The model incorporates the recent CM2 charge scheme that allows as accurate as possible a representation of the electrostatic properties as may be achieved with distributed monopoles. (3) The calculations are parameterized to predict accurate solvation energies on the basis of the unsolvated solute geometry, which makes the calculations not only more affordable but also better suited for playing a role in multi-level treatments. (4) Phosphorus-containing molecules can be handled with significantly more accuracy than solvation models that do not include  $d$  orbitals.

It is well known that DFT is particularly useful for systems containing transition metals, and the creation of the present SM5.42R/DFT model allows us to contemplate extensions to DFT calculations of solvation energies for systems containing transition metals. Such work is in progress.

## VI. SUPPLEMENTARY MATERIAL

Complete tables of calculated  $\Delta G_{EP}$ ,  $G_{CDS}$ ,  $G_S^0$  and experimental  $G_S^0$  for all three parameterizations are given in the supplementary material which is available from Electronic Physics Auxiliary Publication Services (E-PAPS). The E-PAPS consists of one long table of 105 pages.<sup>64</sup>

## ACKNOWLEDGMENTS

This work is in part supported by the National Science Foundation and also by the National Institute of Standards and Technology through the general computational methodology provisions of an Advanced Technology Project sub-contract with Phillips Petroleum Company.

- <sup>1</sup>(a) J. Tomasi and M. Persico, *Chem. Rev.* **94**, 2027 (1994); (b) C. J. Cramer and D. G. Truhlar, in *Reviews in Computational Chemistry*, edited by D. B. Boyd and K. B. Lipkowitz (VCH, New York, 1995), Vol. 6, p. 1; (c) J.-L. Rivail and D. Rinaldi, in *Computational Chemistry: Reviews of Current Trends*, edited by J. Leszczynski (World Scientific, Singapore, 1996), Vol. 1, p. 139.
- <sup>2</sup>(a) K. Heinzinger, in *Supercomputer Simulations in Chemistry*, edited by M. Dupuis (Springer-Verlag, Berlin, 1986), p. 261; (b) W. L. Jorgensen, *Acc. Chem. Res.* **22**, 184 (1989); (c) P. E. Smith and B. Montgomery, *J. Phys. Chem.* **98**, 9700 (1994).
- <sup>3</sup>C. J. Cramer and D. G. Truhlar, *J. Am. Chem. Soc.* **113**, 8305; **113**, 9901 (1991).
- <sup>4</sup>(a) C. J. Cramer and D. G. Truhlar, *Science* **256**, 213 (1992); (b) C. J. Cramer and D. G. Truhlar, *J. Comput.-Aided Molec. Des.* **6**, 629 (1992).
- <sup>5</sup>D. J. Giesen, J. W. Storer, C. J. Cramer, and D. G. Truhlar, *J. Am. Chem. Soc.* **117**, 1057 (1995).
- <sup>6</sup>C. C. Chambers, G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem.* **100**, 16385 (1996).
- <sup>7</sup>D. J. Giesen, M. Z. Gu, C. J. Cramer, and D. G. Truhlar, *J. Org. Chem.* **61**, 8720 (1996).
- <sup>8</sup>D. J. Giesen, G. D. Hawkins, D. A. Liotard, C. J. Cramer, and D. G. Truhlar, *Theor. Chem. Acc.* **98**, 85 (1997).
- <sup>9</sup>G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem.* **100**, 19824 (1996).
- <sup>10</sup>D. J. Giesen, C. C. Chambers, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. B* **101**, 2061 (1997).
- <sup>11</sup>G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. B* **101**, 7147 (1997).
- <sup>12</sup>G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. B* **102**, 3257 (1998).
- <sup>13</sup>G. J. Hoijtink, E. deBoer, P. H. van der Meij, and W. P. Weijland, *Recl. Trav. Chim. Pays-Bas.* **75**, 487 (1956).
- <sup>14</sup>(a) I. Jano, *Compt. Rend. Acad. Sci. (Paris)* **261**, 103 (1965); (b) O. Tapia, in *Quantum Theory of Chemical Reactions*, edited by R. Daudel, A. Pullman, L. Salem, and A. Viellard (Reidel, Dordrecht, 1980), Vol. 2, p. 25.
- <sup>15</sup>W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- <sup>16</sup>R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).
- <sup>17</sup>T. Ziegler and V. Tschinke, in *Density Functional Methods in Chemistry*, edited by J. K. Labanowski and J. W. Andzelm (Springer-Verlag, New York, 1991), p. 139.
- <sup>18</sup>*Density Functional Theory of Molecules, Clusters, and Solids*, edited by D. E. Ellis (Kluwer Academic, Dordrecht, 1995).
- <sup>19</sup>P. Margl, J. C. W. Lohrenz, T. Ziegler, and P. E. Blöchl, *J. Am. Chem. Soc.* **118**, 4434 (1996), and references therein.
- <sup>20</sup>J. L. Chen, L. Noodleman, D. A. Case, and D. Bashford, *J. Phys. Chem.* **98**, 11059 (1994).
- <sup>21</sup>C. Adamo and F. Leij, *Chem. Phys. Lett.* **223**, 54 (1994).
- <sup>22</sup>A. Fortunelli and J. Tomasi, *Chem. Phys. Lett.* **231**, 34 (1994).
- <sup>23</sup>(a) T. N. Truong and E. V. Stefanovich, *Chem. Phys. Lett.* **240**, 253 (1995); (b) T. N. Truong and E. V. Stefanovich, *J. Chem. Phys.* **103**, 3709 (1995); (c) E. V. Stefanovich and T. N. Truong, *Chem. Phys. Lett.* **244**, 65 (1995).
- <sup>24</sup>J. Andzelm and C. Kolmel, *J. Chem. Phys.* **103**, 9312 (1995).
- <sup>25</sup>G. J. Tawa, R. L. Martin, L. R. Pratt, and T. V. Russo, *J. Phys. Chem.* **100**, 1515 (1996).
- <sup>26</sup>J. Li, T. Zhu, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. A* **102**, 1820 (1998).
- <sup>27</sup>R. S. Mulliken, *J. Chem. Phys.* **23**, 1833 (1955).
- <sup>28</sup>P.-O. Löwdin, *J. Chem. Phys.* **18**, 365 (1950).
- <sup>29</sup>J. W. Storer, D. J. Giesen, C. J. Cramer, and D. G. Truhlar, *J. Comput.-Aided Molec. Des.* **9**, 87 (1995).
- <sup>30</sup>(a) A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988); (b) J. P. Perdew, in *Electronic Structure of Solids*, edited by P. Ziesche and H. Eschrig (Akademie Verlag, Berlin, 1991), p. 11; (c) J. P. Perdew, K. Burke, and M. Ernzerhof, *ACS Symp. Ser.* **629**, 453 (1996); (d) J. P. Perdew, K. Burke, and Y. Wang, *Phys. Rev. B* **54**, 16533 (1996).
- <sup>31</sup>(a) R. E. Easton, D. J. Giesen, A. Welch, C. J. Cramer, and D. G. Truhlar, *Theor. Chim. Acta* **93**, 281 (1996); (b) J. Li, C. J. Cramer, and D. G. Truhlar, *Theor. Chem. Acc.* **99**, 192 (1998).
- <sup>32</sup>N. Godbout, D. R. Salahub, J. Andzelm, and E. Wimmer, *Can. J. Chem.* **70**, 560 (1992).
- <sup>33</sup>W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.* **56**, 2257 (1972).
- <sup>34</sup>P. C. Hairaran and J. A. Pople, *Theor. Chim. Acta* **28**, 213 (1973).
- <sup>35</sup>M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, *J. Chem. Phys.* **77**, 3654 (1982).
- <sup>36</sup>The 6-31G\* basis for Br is taken from Ref. 61.
- <sup>37</sup>T. Kozaki, K. Morihashi, and O. Kikuchi, *J. Am. Chem. Soc.* **111**, 1547 (1989).
- <sup>38</sup>O. Kikuchi, T. Matsvoka, H. Sawahata, and O. Takahashi, *J. Mol. Struct.: THEOCHEM* **305**, 79 (1994).
- <sup>39</sup>W. C. Still, A. Tempczak, R. C. Hawley, and T. Hendrickson, *J. Am. Chem. Soc.* **112**, 6127 (1991).
- <sup>40</sup>I. Mayer, *Chem. Phys. Lett.* **97**, 270 (1983).
- <sup>41</sup>A. Bondi, *J. Phys. Chem.* **68**, 441 (1964).
- <sup>42</sup>*CRC Handbook of Chemistry and Physics*, 75th ed., edited by D. R. Lide (CRC Press, Boca Raton, FL, 1995).
- <sup>43</sup>M. H. Abraham, *Chem. Soc. Rev.* **22**, 73 (1993).
- <sup>44</sup>M. H. Abraham, *J. Phys. Org. Chem.* **6**, 660 (1993).
- <sup>45</sup>M. H. Abraham, H. S. Chadha, G. S. Whiting, and R. C. Mitchell, *J. Pharm. Sci.* **83**, 1085 (1994).
- <sup>46</sup>D. Suleiman and C. A. Eckert, *J. Chem. Eng. Data* **39**, 692 (1994).
- <sup>47</sup>S. Cabani, P. Gianni, V. Mollica, and L. J. Lepori, *Solution Chem.* **10**, 563 (1981).
- <sup>48</sup>M. H. Abraham, G. S. Whiting, R. Fuchs, and E. J. Chambers, *J. Chem. Soc., Perkin Trans. 2* **1990**, 291.
- <sup>49</sup>J. Hine and P. K. Mookerjee, *J. Org. Chem.* **40**, 287 (1975).
- <sup>50</sup>D. D. Wagman, *J. Phys. Chem. Ref. Data Suppl.* **2**, 11 (1982).
- <sup>51</sup>R. Wolfenden, *Biochemistry* **17**, 201 (1978).
- <sup>52</sup>R. D. Wauchope and R. Haque, *Can. J. Chem.* **50**, 133 (1972).
- <sup>53</sup>P. Han and D. M. Bartles, *J. Phys. Chem.* **94**, 7294 (1990).
- <sup>54</sup>A. J. Leo, *Masterfile* from MedChem Software, BioByte Corp., P. O. 517, Claremont, CA 91711-0157 (1994).
- <sup>55</sup>D. M. Ferguson, D. A. Pearlman, W. C. Swope, and P. A. Kollman, *J. Comput. Chem.* **13**, 362 (1992).
- <sup>56</sup>R. G. Pearson, *J. Am. Chem. Soc.* **108**, 6109 (1986).
- <sup>57</sup>J. Florian and A. Warshel, *J. Phys. Chem. B* **101**, 5583 (1997).
- <sup>58</sup>J. Andzelm and E. Wimmer, *J. Chem. Phys.* **96**, 1280 (1992).
- <sup>59</sup>DGAUSS 4.0, Oxford Molecular, Ltd. (1996).
- <sup>60</sup>AMSOL version 6.5, G. D. Hawkins, D. G. Giesen, C. C. Chambers, G. C. Lynch, I. Rossi, J. W. Storer, D. Rinaldi, D. A. Liotard, C. J. Cramer, and D. G. Truhlar, University of Minnesota, Minneapolis, 1998.
- <sup>61</sup>GAUSSIAN 94, M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Peterson, J. A. Montgomery, K. Raghavachari, M. A. A.-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. DeFrees, J. Baker, J. J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh, 1995.
- <sup>62</sup>MN-GSM version 98.2.3, J. Li, G. D. Hawkins, D. A. Liotard, C. J. Cramer, and D. G. Truhlar, University of Minnesota, Minneapolis, 1998.
- <sup>63</sup>M. J. S. Dewar, E. G. Zebisch, and J. J. P. Stewart, *J. Am. Chem. Soc.* **107**, 3902 (1985).
- <sup>64</sup>See AIP Document No. E-PAPS: E-JCPSA 6-109-508844 for a table of solvation free energies and terms in the solvation free energy. E-PAPS document files may be obtained free of charge from the FTP server at <http://www.aip.org/epaps/epaps.html> or from [ftp.aip.org](ftp://ftp.aip.org) in the directory /epaps/. For further information, send e-mail to [paps@aip.org](mailto:paps@aip.org) or fax to (516)576-2223.