JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

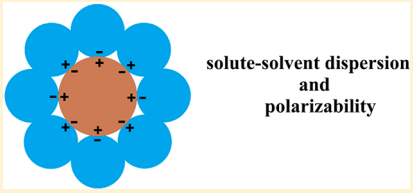
Uniform Treatment of Solute−Solvent Dispersion in the Ground and Excited Electronic States of the Solute Based on a Solvation Model with State-Specific Polarizability

Aleksandr V. Marenich, Christopher J. Cramer,* and Donald G. Truhlar*

Department of Chemistry, Chemical Theory Center, and Supercomputing Institute, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, Minnesota 55455-0431, United States

Supporting Information

ABSTRACT: We present a new kind of treatment of the solute−solvent dispersion contribution to the free energy of solvation using a solvation model with state-specific polarizability (SMSSP). To evaluate the solute−solvent dispersion contribution, the SMSSP model utilizes only two descriptors, namely, the spherically averaged dipole polarizability of the solute molecule (either in its ground or excited electronic state) and the refractive index of the solvent. The model was parametrized over 643 ground-state solvation free energy data for 231 solutes in 14 nonpolar, non-hydrogen-bonding solvents. We show that the SMSSP model is applicable to solutes in both the ground and the excited electronic state. For example, in comparison to available experimental data, the model yields qualitatively accurate predictions of the solvatochromic shifts for a number of systems where solute−solvent dispersion is the dominant contributor to the shift.



1. INTRODUCTION

For computational convenience, solute−solvent interactions can be partitioned into (i) the bulk electrostatic contribution associated with the polarization free energy of the solute−solvent system when the solute is inserted (as well as with the change in the solute's internal electronic energy upon transferring from the gas phase to the liquid phase) and (ii) everything else. The second component is called the non-bulk-electrostatic contribution (that is, everything except bulk electrostatics), and it is associated with cavitation, dispersion, and solvent-structure effects (CDS), where the solvent-structure contributions include effects such as hydrogen bonding solvent structure breaking, exchange repulsion, solute−solvent charge transfer, dielectric saturation, and electrostriction effects.[1] The physics of these effects is mainly localized in the first solvation shell. (We use the terms "solvation sphere" and "solvation shell" interchangeably.)

The partition into i and ii is not unique since it depends on the definition of the solute atomic radii[2−4] for the bulk electrostatic calculation. The partition of component ii is even more problematic since even in an isolated diatomic molecule one can separate contributions like damped dispersion and exchange repulsion only approximately,[5] and some solvation models combine non-bulk-electrostatic interactions such as exchange repulsion with the dispersion energy into a single component called the van der Waals contribution to the interaction energy.[6,7] Note that dispersion-like interactions (i.e., London forces[8]) are damped[9,10] from their asymptotic form at van der Waals distances characteristic of solute interactions with the first solvation sphere, and they include overlap effects and more general forms of medium-range correlation energy. Nevertheless, the approximate physical separation of contributions is widely used in solvation modeling and is often a key ingredient in its success. With these understandings, in the present article we will continue to refer to the electron correlation contribution to the solute−solvent interaction energy in the first solvation sphere as dispersion.

By invoking the dielectric continuum approximation to describe the solvent, the bulk electrostatic contribution to the solute−solvent interaction energy can be understood as the interaction between a polarized solute molecule and a mean reaction field of solvent molecules self-consistently polarized by the solute, and the interaction is a long-range one as it falls off as $1/R^4$ (the same as ion-induced dipole induction interactions between particles, except that with particles R is an interparticle distance, but here R is the distance between a charge center in the solute and an arbitrary point in the continuum dielectric medium representing the solvent). To calculate this polarization interaction, one neglects complications due to the first solvation shell where electron exchange is important and approximates the wave function of a system (Ψ_{MS}) containing a solute molecule (M) and a collection of solvent molecules (S) by a Hartree product of Ψ_M and Ψ_S. Thus, not only electron exchange but also electron correlation between the electrons in the solute molecule and the solvent's electrons is neglected in calculating the bulk polarization. However, at shorter distances (comparable with the size of the first solvation shell) the solute−solvent electron correlation effects become more important, resulting in attractive interactions between mutually induced dipoles and higher multipoles; these are the solute−solvent dispersion interactions.

Received: April 19, 2013

dx.doi.org/10.1021/ct400329u | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

The dispersion interaction between two induced dipoles varies asymptotically with the distance R as $1/R^6$. The shorter-range solute−solvent interactions also include exchange repulsion between two systems due to the Pauli exclusion principle.[1] Another first-shell effect is cavitation. The cavitation energy is sometimes defined as the work spent in forming a solute cavity of appropriate shape and volume inside the solvent in the absence of solute−solvent interactions, assuming that it is proportional to the macroscopic or microscopic surface tension of the solvent and the solvent accessible surface area of the solute molecule.[6] In general, dispersion and repulsion effects are inseparable from cavitation and other first-shell effects including specific solvent-structure effects (such as induced changes in hydrogen bonding between solvent molecules). For example, if one makes a cavity in a solvent to accommodate a solute molecule, the solvent molecules in the first solvation shell will have solute−solvent dispersion interaction and exchange repulsive interactions at the expense of solvent−solvent ones, and such a trade-off can significantly impact the solvent's structural properties.[1]

Some portion of non-bulk-electrostatic effects can be recovered using the dielectric continuum approach by making the solute−solvent dielectric boundary dependent on solvent properties, e.g., Abraham's hydrogen bond acidity and basicity parameters[11] to account for contributions due to hydrogen bonding effects and solvent-structure-breaking effects, and/or by using mixed discrete−continuum solvation models when solute−solvent hydrogen bonding and charge transfer effects are treated explicitly. In general, non-bulk-electrostatic effects (including cavitation, dispersion, and repulsion) can be treated by using various semiempirical functional forms combined with bulk electrostatics. Our previous studies employed empirical atomic surface tension terms[12−14] called CDS terms that account for cavity formation (C), dispersion (D), and solvent-structure (S) effects. In another approach, Floris et al.[15,16] derived the dispersion−repulsion term from classical pair potentials; this approach is available in *Gaussian 03*[17] and *Gaussian 09*,[18] along with a calculation of the cavitation energy based on the scaled particle theory of Pierotti,[19] as modified by Claverie.[20] Other treatments of non-bulk-electrostatic effects are presented in refs 6 and 21−28 (see also references therein).

Here, we present a simple way to treat the solute−solvent dispersion contribution to the free energy of solvation. The new method uses a solvation model based on the spherically averaged molecular polarizability of the solute molecule. Importantly, it will be shown that the new model, to be called solvation model based on state-specific polarizability (SMSSP), can be successfully applied to solutes in both ground and excited electronic states. It will be shown that, within the SMSSP approach, the solute−solvent dispersion contribution can be treated simply as a polarizability-dependent semi-empirical correction that does not involve a modification of the solute's electronic wave function. To study the solute−solvent dispersion effects on the solute charge distribution in either the ground or excited electronic state and related molecular properties, one would need to use alternative approaches (see, for example, refs 29 and 30) that utilize a dispersion operator or dispersion−repulsion operator explicitly included in the solute Hamiltonian.

2. THEORY

In our earlier work,[14] the fixed-concentration free energy of solvation is defined as a free energy of transfer of a solute from an ideal gas at a concentration of 1 mol/L to a 1 M ideal solution according to

$$\Delta G_S^* = \Delta G_{EP} + G_{CDS} + \Delta G_N \quad (1)$$

where the bulk electrostatic component ΔG_{EP} is a sum of the change in the solute's internal electronic energy (ΔE_E) in transferring from the gas phase to the liquid phase at the same geometry and the polarization free energy (G_P) of the solute−solvent system when the solute is inserted (including both the favorable interaction free energy and the unfavorable work of polarizing the solvent that cancels half of the favorable interaction in the linear response regime[1]); G_{CDS} is the component of the free energy that is nominally associated with cavitation, dispersion, and solvent structure. Formally we define ΔG_{EP} and G_{CDS} with the solute geometry fixed at the gas-phase equilibrium structure, and ΔG_N is the change in ΔG_S^* due to a change in nuclear coordinates. Note that standard-state solvation free energies are related to fixed-concentration ones as

$$\Delta G_S^o = \Delta G_S^* + \Delta G_{conc}^{*\to o} \quad (2)$$

where $\Delta G_{conc}^{*\to o}$ is due to the difference in concentrations, if any, in the gas-phase standard state from the solution-phase one. If the gas-phase standard state is an ideal gas at a pressure of one bar and the liquid-phase standard state is an ideal solution with a concentration of 1 M, the quantity $\Delta G_{conc}^{*\to o}$ is equal to 1.9 kcal/mol.

In the present article, we define a new solvation model for systems where solute−solvent hydrogen bonding is not present (only such systems are considered in the present paper); we treat G_{CDS} as a sum of G_D and G_{CR} where G_D is an attractive interaction term associated with solute−solvent dispersion and G_{CR} is a potential associated with cavitation and with repulsive forces between the solute and the solvent's molecules (mainly due to exchange repulsion). The new model is called solvation model based on state-specific polarizability, abbreviated SMSSP.

Since all calculations of ΔG_S^* in the present work are based on rigid, gas-phase geometries, the ΔG_N component in eq 1 is taken to be zero. We have illustrated previously that for small, rigid molecules like those in most training sets, this approximation is reasonable.[31] In that case, eq 1 can be rewritten as

$$\Delta G_S^* = \Delta G_{EP} + G_D + G_{CR} \quad (3)$$

The dispersion contribution is treated by the following formula

$$G_D = \sigma_D \alpha_M \frac{n_S^2 - 1}{n_S^2 + 2} \quad (4)$$

where σ_D is an empirical constant with units of energy per unit volume, α_M is the spherically averaged molecular polarizability of the solute molecule M, and n_S is the refractive index of the solvent S at optical frequencies at 293 K. This formula is motivated in the Appendix.

The cavitation−repulsion term G_{CR} is treated by the following equation:

$$G_{CR} = S_{CR} \gamma_S A_M \quad (5)$$

where S_{CR} is a unitless empirical constant, γ_S is the macroscopic surface tension of the solvent S at the air/solvent interface at 298 K, and A_M is the total solvent accessible surface area

Table 1. Solvents Used in the SMSSP Training Set

	solvent descriptors[a]				solvent descriptors[a]		
solvent	ε	n	γ	solvent	ε	n	γ
CCl_4[b]	2.228	1.460	38.0	isooctane	1.936	1.392	26.4
cyclohexane	2.017	1.427	35.5	nonane	1.961	1.405	32.2
decalin[c]	2.196	1.453	43.8	pentadecane	2.033	1.432	38.3
decane	1.985	1.410	33.6	pentane	1.837	1.358	22.3
dodecane	2.006	1.422	35.9	perfluorobenzene	2.029	1.378	31.7
heptane	1.911	1.388	28.3	tetrachloroethene	2.268	1.505	45.2
hexadecane	2.040	1.435	38.9	undecane	1.991	1.440	34.9

[a]Dielectric constants (ε) and refractive indexes (n) are dimensionless; surface tension coefficients (γ) are in units of cal mol^{-1} Å^{-2} (note that 1 dyn/cm =1.44 cal mol^{-1} Å^{-2}). [b]Carbon tetrachloride. [c]Cis/trans mixture.

(SASA) of the solute molecule M calculated by summing over accessible surface areas of individual atoms in the solute molecule calculated by the analytic surface area (ASA) algorithm[32] using Bondi's values[33] for van der Waals atomic radii. Note that for atoms where Bondi does not give a value, we recommend the values of Mantina et al.,[34,35] which were specifically developed to be consistent with Bondi's values. We further offset the solute−solvent boundary by using a universal effective solvent-shell radius of 0.4 Å, as in earlier work.[14] (This is equivalent to adding 0.4 Å to all atomic radii.)

We caution the reader to be careful not to confuse the different kinds of radii. The atomic radii used for bulk electrostatics are called intrinsic Coulomb radii. The atomic radii used for non-bulk-electrostatic contributions to the solvation free energies are called van der Waals radii, and the offsets used to convert van der Waals surfaces to solvent-accessible surfaces (that model a surface passing through the center of the first solvation shell) are called effective solvent radii.

The SMSSP model has been parametrized by simultaneously optimizing the two model parameters (σ_D and S_{CR}) using a set of 643 experimental reference solvation free energies (taken from the Minnesota Solvation Database[36]) for 231 neutral solute molecules in their ground electronic states in 14 nonpolar solvents that have zero values for both of Abraham's hydrogen bond acidity and basicity parameters.[11] These solvents are listed in Table 1 along with the corresponding values of the solvent descriptors used in SMSSP, namely, bulk dielectric constants (ε), refractive indexes (n), and macroscopic surface tensions (γ); these solvent descriptors are all obtained from the Minnesota Solvent Descriptor Database.[37] It is reasonable to assume that the non-bulk-electrostatic contributions to the free energy of solvation in the selected solvents will be dominated by the solute−solvent dispersion interaction, and the solvent-structure effects due to hydrogen bonding can be safely neglected in the SMSSP parametrization in these solvents. The solute molecules used to obtain the two parameters contain at most H, C, N, O, F Si, P, S, Cl, Br, and I. However, because the model involves only two very general parameters (σ_D and S_{CR}), we believe it can be applied to any solute, even those containing other elements. The experimental reference data used in the SMSSP parametrization are provided in the Supporting Information (SI).

The optimization of σ_D and S_{CR} was carried out by minimizing the following error function:

$$\chi = \sum_{I=1}^{643} [\Delta G_S^*(\text{expt.}, I) - \Delta G_{EP}(I) - G_D(I) - G_{CR}(I)]^2 \quad (6)$$

where the index I runs over 643 experimental (expt.) solvation energies for neutral solutes in 14 solvents, $\Delta G_{EP}(I)$ is the bulk-electrostatic polarization contribution from a self-consistent reaction field calculation, $G_D(I)$ is defined by eq 4, and $G_{CR}(I)$ is defined by eq 5.

3. COMPUTATIONAL DETAILS

3.1. SMSSP. The quantity $\Delta G_{EP}(I)$ in eq 6 is calculated by solving the nonhomogeneous-dielectric Poisson equation (NPE)[7] for bulk electrostatics based on the continuous charge density of the solute molecule by using the Integral Equation Formalism Polarizable Continuum Model (IEFPCM) algorithm.[38] In the SMSSP model, we use the SMD[13] intrinsic Coulomb radii for the bulk electrostatic calculations. We will label this treatment of bulk electrostatics as NPE(D). We average the ΔG_{EP} value over calculations with five density functionals as

$$\Delta G_{EP}(I) = \frac{1}{5}\sum_{k=1}^{5} \Delta G_{EP}(I, J_k) \quad (7)$$

where k runs over B3LYP,[39−42] mPW1PW,[43] M06-L,[44] M06,[45,46] and M06-2X,[45,46] all employed with the MG3S[47] basis set. Averaging the ΔG_{EP} value over several different methods helps eliminate idiosyncratic errors in the description of bulk electrostatics associated with individual methods. The bulk electrostatic calculations were performed using *Gaussian 09*.[18]

3.2. Nonstandard Versions of SMSSP. We have also tested using the universal force field (UFF) radii[48] scaled by 1.1 (this is *Gaussian 09*'s default for bulk electrostatics) in the solution of the nonhomogeneous-dielectric Poisson equation, and we will call this approach NPE(1.1×UFF).

We have also tested using the new polarizability−cavitation terms of the SMSSP model in conjunction with bulk electrostatics computed by the SM12 solvation model;[14] this method for bulk electrostatics utilizes the generalized Born (GB) approximation[49,50] based on partial atomic charges and is here labeled as GB. In the GB calculations, we employed ChElPG charges[51] and the intrinsic Coulomb radii used by the SM12 model, and these calculations were carried out by using a locally modified version of *Gaussian 09* called Minnesota Gaussian Solvation Module[52] or *MN-GSM*.

In some calculations, we have also tested using the 6-31G(d) basis set[53] instead of MG3S.

dx.doi.org/10.1021/ct400329u | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. Errors in Ground-State Free Energies of Solvation (in kcal/mol) Computed Using SMSSP and Other Models[a]

NBE model[b]	BE model[c]	xcF	basis	MSE	MUE	notes
SMSSP	default	B3LYP	MG3S	−0.04	0.46	*d*
SMSSP	default	mPW1PW	MG3S	−0.10	0.48	*d*
SMSSP	default	M06-L	MG3S	0.14	0.47	*d*
SMSSP	default	M06	MG3S	−0.02	0.46	*d*
SMSSP	default	M06-2X	MG3S	−0.12	0.48	*d*
SMSSP	default	B3LYP	6-31G(d)	0.15	0.49	*d*
SMSSP	NPE(1.1 × UFF)	B3LYP	MG3S	0.49	0.63	*e*
SMSSP	GB	mPW1PW	MG3S	0.33	0.68	*f*
SMD	default	B3LYP	MG3S	0.11	0.57	*g*
SMD	default	mPW1PW	MG3S	0.06	0.56	*g*
SMD	default	M06-L	MG3S	0.30	0.59	*g*
SMD	default	M06	MG3S	0.14	0.57	*g*
SMD	default	M06-2X	MG3S	0.03	0.56	*g*
SMD	default	B3LYP	6-31G(d)	0.31	0.57	*g*
SM12	default	mPW1PW	MG3S	0.04	0.54	*h*
none	NPE(1.1 × UFF)	B3LYP	MG3S	3.26	3.29	*i*
none	NPE(D)	B3LYP	MG3S	2.74	2.80	*j*
standard PCM	NPE(1.1 × UFF)	B3LYP	MG3S	8.04	8.04	*k*
standard PCM	NPE(D)	B3LYP	MG3S	7.46	7.46	*l*
PCMopt	NPE(D)	B3LYP	MG3S	−0.08	0.81	*m*

[a]NBE refers to a model for the non-bulk-electrostatic contribution to the free energy of solvation. BE refers to a model for bulk electrostatics (GB or NPE), and the next four columns show the exchange-correlation functional (xcF), basis set, mean signed error, and mean unsigned error. Calculations were performed over 240 solute data for the last three rows and over 643 data for the other cases. [b]The SMSSP NBE component was calculated using σ_D and S_{CR}. The SMD and SM12 NBE components were calculated using the default parameters for these models. The standard PCM calculations are explained in section 3.3. The PCMopt calculations are explained in section 3.4. [c]The default BE model for SMSSP and SMD is the NPE solver with the SMD radii [NPE(D)]. We use the IEFPCM algorithm as the NPE solver. The default BE model for SM12 is GB. [d]An example of a standard SMSSP calculation (recommended method). [e]A nonstandard calculation to show the effect of changing the BE model. [f]An example of an SMSSP(GB) calculation. [g]An example of a standard SMD calculation. [h]An example of a standard SM12 calculation. [i]Only bulk electrostatics (*Gaussian 09* default). [j]Only bulk electrostatics (for comparison, the intrinsic Coulomb radii are changed as compared to the previous row). [k]Standard PCM treatment of cavitation, dispersion, and repulsion (see section 3.3). [l]Variation of previous row to show effect of changing bulk electrostatics. [m]Optimized PCM treatment of cavitation, dispersion, and repulsion (see section 3.4).

3.3. Standard PCM. The SMSSP results were also compared to those obtained using an alternative approach to calculate cavitation, dispersion, and repulsion energies. In particular, we use the polarizable continuum model (PCM)[6,7] that includes the treatment of cavitation, dispersion, and repulsion in addition to the treatment of bulk electrostatics. In this model, G_{CDS} of eq 1 is replaced by

$$G_D + G_R + G_C = G^{FT} + G^{PC} \quad (8)$$

where the terms on the left account respectively for dispersion, repulsion, and cavitation, and the right-hand side indicates that these terms are modeled by combining the model of Floris and Tomasi (FT)[15,16] to calculate G_D and G_R based on a dispersion−repulsion term derived from classical pair potentials (see, for example, eq 20 in ref 13) with the scaled particle theory of Pierotti,[19] as modified by Claverie[20] (here abbreviated PC; see eq 21 in ref 13) for the cavitation. Note that, as illustrated in the equations just cited, the FT calculations depend on a choice of van der Waals atomic radii of all atoms in the solute, and the PC calculations involve an effective solvent radius both separately and added to the van der Waals atomic radii.

The FT and PC calculations were carried out using *Gaussian 09*.[18] Equation 8 was calculated by default in a *Gaussian 03* PCM calculation,[17] but in *Gaussian 09* one needs to request the calculation of eq 8 by using the keywords *cav*, *dis*, *rep*, and *IOp*(5/33=2).

The FT and PC calculations were performed for 240 solute data in the three solvents (namely, in CCl_4, cyclohexane, and *n*-heptane) for which the corresponding FT and PC parameters are available (FT and PC parameters are not available for all 14 solvents studied in the present work).

In the FT calculations, we used the *Gaussian 09* default values for the van der Waals atomic radii of the solute atoms; in particular, we used the UFF atomic radii.[48] As will be shown later, calculations based on these radii can reproduce the results obtained by Floris and Tomasi[15] reasonably well. For example, the van der Waals atomic radii equal 1.443 and 1.926 Å, respectively, for H and C. In the PC calculations, we used the default UFF atomic radii, and for the effective solvent radius we used the *Gaussian 09* default values. In particular, we used the values of 2.685, 2.815, and 3.125 Å for the effective solvent radii of CCl_4, cyclohexane, and heptane, respectively. For example, in the case of CCl_4, the sum of the van der Waals atomic radii and the effective solvent radius equals 4.128 and 4.611 Å, respectively, for H and C.

The non-bulk-electrostatic energy from the FT and PC computations were combined with the bulk electrostatic energy obtained by NPE(D) or NPE(1.1×UFF) to produce free energies of solvation for comparison to the experiment.

3.4. Optimized PCM. To minimize the deviation between computed and reference solvation free energies over the set of 240 solvation data used in FT and PC calculations, we approximately optimized the models, leading to the following choices of parameters. For the FT calculations, we use the Bondi van der Waals radii plus 1.3 Å (there is no effective solvent radius in the FT calculations). For example, with this choice, the van der Waals atomic radii on H and C equal 2.5

dx.doi.org/10.1021/ct400329u | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

and 3.0 Å, respectively. For the PC calculations we took the van der Waals atomic radii as the Bondi van der Waals radii plus 1.3 Å, and we took the effective solvent radius to be 1.3 Å for all solvents. For example, with this choice, when the calculation calls for the sum of the van der Waals atomic radii plus the effective solvent radius, we use 3.8 Å and 4.3 Å, respectively, for H and C. The resulting PCM model based on the FT and PC computations that utilize these approximately optimized radii will be called PCMopt.

Examples of *Gaussian 09* input files for FT and PC calculations with the PCMopt parameters are provided in the SI.

3.5. Geometries and Polarizabilities. In all calculations we use ground-state molecular geometries optimized for the gas phase by using the M06-2X exchange-correlation functional[45,46] with the MG3S[47] basis set.

The SMSSP solvation model is based on gas-phase molecular polarizabilities that can be obtained analytically if analytical second-order energy derivatives are available, by numerical differentiation of analytical gradients when analytical gradients but not Hessians are available, or by double numerical differentiation when only energies are available. The latter case is least efficient and requires in general up to seven single-point energy calculations (one at zero electric field and two per each direction x, y, and z).

The spherically averaged polarizability used in eq 4 for each tested solute molecule is calculated as the average over five values computed analytically in the gas phase by *Gaussian 09* using the exchange−correlation functionals B3LYP,[39−42] mPW1PW,[43] M06-L,[44] M06,[45,46] and M06-2X[45,46] with the MG3[47,54] basis set. These α_M values (which are given in the SI) are expected to have a mean reliability of ∼0.7 Å^3 (or 10%), as estimated using experimental reference values of α_M for several dozen solute molecules of our training set for which experimental polarizabilities are available in the reference literature.[55]

4. RESULTS AND DISCUSSION

4.1. Ground States. The SMSSP parameters σ_D and S_{CR} were optimized over 643 reference solvation data for 231 unique solutes in 14 solvents (i.e., having 115 solutes per parameter). By minimizing the error function in eq 6, we obtain the following values of the SMSSP model parameters: σ_D = −1.850 kcal mol^{-1} Å^{-3} and S_{CR} = 0.369 (S_{CR} is dimensionless). These parameters were obtained using the ΔG_{EP} values averaged over the five NPE(D)/X/MG3S[47] methods for bulk electrostatic calculations (where X denotes a choice of exchange−correlation functional, in particular, B3LYP,[39−42] mPW1PW,[43] M06-L,[44] M06,[45,46] and M06-2X[45,46]). The SMSSP parameters σ_D and S_{CR} were optimized over 643 reference solvation data for 231 unique solutes in 14 solvents (i.e., having 115 solutes per parameter).

To test the robustness of the fit we employed the following cross-validation procedure. We removed every second datum out of 643 data arranged as listed in the SI, and we reoptimized the model using the remaining 322 data. We found that after exclusion of even 50% of the data from the optimization, the reoptimized values of σ_D and S_{CR} change only by 4% (σ_D) and 8% (S_{CR}). We have also tested using more data in the SMSSP parametrization by adding 97 reference free energies of solvation for 67 solutes in two additional solvents, namely, *n*-hexane and *n*-octane.[36] The optimized values of σ_D and S_{CR}

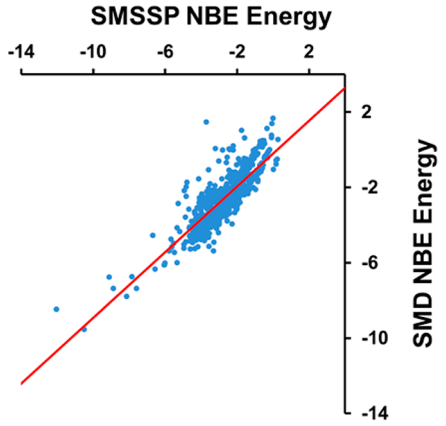


Figure 1. The non-bulk-electrostatic (NBE) contribution to the SMSSP free energy of solvation over 643 solvation data in 14 nonpolar solvents (in kcal/mol) versus the corresponding NBE contribution to the SMD free energy of solvation determined by the SMD CDS term. The mean signed and the mean unsigned deviation (relative to SMSSP NBE) are 0.16 and 0.59 kcal/mol, respectively. The value of R^2 is 0.72 where R is a correlation coefficient.

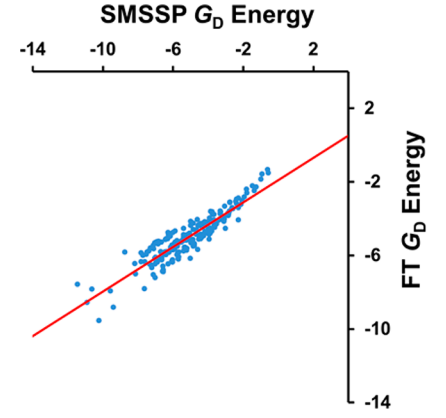


Figure 2. The solute−solvent dispersion (G_D) to the SMSSP free energy of solvation over 240 solvation data in three nonpolar solvents (in kcal/mol) versus the corresponding dispersion contribution obtained from PCMopt FT computations. The mean signed and the mean unsigned deviation (relative to SMSSP) is 0.14 and 0.70 kcal/mol, respectively. The value of R^2 is 0.83 where R is a correlation coefficient.

Table 3. Solute−Solvent Dispersion Energies (G_D) for Selected Solutes by Various Methods in Comparison with the Reference Solvation Free Energies (ΔG_S*; in kcal/mol)

quantity	model[a]	solvent	C_3H_8	C_5H_{12}	C_8H_{18}
G_D	SMSSP	cyclohexane	−2.8	−4.5	−7.1
G_D	PCMopt	cyclohexane	−3.6	−5.0	−7.0
G_D	standard PCM	cyclohexane	−8.2	−11.4	−16.1
ΔG_S*	exptl.[b]	cyclohexane	−2.1	−3.5	−5.6
G_D	SMSSP	water	−2.2	−3.6	−5.7
G_D	PCMopt	water	−3.8	−5.2	−7.4
G_D	standard PCM	water	−8.9	−12.3	−17.4
G_D	Floris, Tomasi[c]	water	−8.7	−12.3	−17.2
ΔG_S*	exptl.[b]	water	2.0	2.3	2.9

[a]SMSSP refers to the SMSSP dispersion energy calculated by eq 4 for given solutes in cyclohexane (n = 1.427) and in water (n = 1.333). See sections 3.3 and 3.4 for further details. [b]Reference 36. [c]From Table 1/Model I in ref 15.

dx.doi.org/10.1021/ct400329u | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 4. Gas-Phase Polarizabilities (α, in Å^3) and Vertical Excitation Energies (ω, in cm^{-1}) by Various Methods, and Dispersion Solvatochromic Shifts ($\Delta\omega_D$, in cm^{-1}) Estimated by eq 10[a]

			$\alpha_{ES} - \alpha_{GS}$		ω		$\Delta\omega_D$	
molecule	method	α_{GS}	L_b	L_a	L_b	L_a	L_b	L_a
benzene		$^1A_{1g}$	$^1B_{2u}$	$^1B_{1u}$				
	M06-2X, TDDFT	9.8	0.8	3.3	45 190	51 283	127	542
	M06, TDDFT	10.0	1.3	3.9	42 883	47 050	215	654
	CCSD, EOM-CCSD	9.6	0.2	4.2	42 389	52 904	31	699
	theory[b]	8.8	0.1	7.7	40 166	59 443	18	1275
	theory[c]	10.2	0.6	6.2	41 779	52 668	98	1041
	theory[d]				40 973	52 749		
	exptl.[e]	10.2	0.9				148	
	exptl.[f]				39 533	49 970		
azulene		$^1A_1$	$^1B_2$	$^1A_1$				
	M06-2X, TDDFT	18.8	−0.7	1.4	19 391	31 122	−123	234
	M06, TDDFT	19.2	−0.4	1.0	19 027	28 377	−64	172
	CCSD, EOM-CCSD	19.0	−1.3	2.4	18 596	32 310	−220	403
	theory[g]				18 551	27 907		
	exptl.[h]				14 270	29 760		
naphthalene		^1A_g	$^1B_{2u}$	$^1B_{1u}$				
	M06-2X, TDDFT	17.0	1.8	8.1	37 950	38 473	307	1344
	M06, TDDFT	17.5	2.7	8.3	35 709	34 918	452	1387
	CCSD, EOM-CCSD	15.8	1.2	10.5	36 236	41 389	202	1753
	theory[b]	15.0	2.4	14.5	34 762	46 054	403	2422
	theory[d]				34 198	38 473		
	exptl.[e]	17.4	1.6 ± 0.1	9.8 ± 1.6[i]			267	1627
	exptl.[h]				32 456	36 417		
anthracene		^1A_g	$^1B_{2u}$	$^1B_{1u}$				
	M06-2X, TDDFT	25.8	3.5	13.0	33 194	29 187	575	2166
	exptl.[i]	25.3[j]		17 ± 2				2776
	exptl.[k]					27 643		
	exptl.[l]				28 055	27 695		
coronene		$^1A_{1g}$	$^1B_{2u}$	$^1B_{1u}$				
	M06-2X, TDDFT	43.1	5.2	11.8	28 664	31 261	870	1971
	estimated[m]	42.5						
	exptl.[k]				24 067	29 089		

[a]The table presents the data for the ground electronic state (GS) and for the two lowest excited (spin-singlet) electronic states denoted as L_a and L_b using Platt's notations (see ref 58). Calculations were performed using M06-2X, M06, or CCSD for the ground state and TDM06-2X, TDM06, or EOM-CCSD for excited-state calculations, with the MG3 basis set. [b]CASSCF calculation with a DZ-quality basis set from ref 67 noted as CAS1/B2 in ref 67. [c]Coupled cluster theory calculation with an extended Sadlej basis set from ref 68 noted as CCSD/"Sadlej+6" in ref 68. [d]This is the best estimate from ref 69 based on CC3 calculations with large basis sets for benzene and based on CASPT2/TZVP calculations for naphthalene. [e]Gas-phase polarizabilities determined from static Stark effect measurements in refs 70 and 71, unless noted otherwise. [f]Values from Table 1 (band c) and Table 2 (band b) of ref 72 for the gas-phase UV absorption spectrum of benzene (see also refs 73 and 74). [g]TDDFT(SOAP) calculation with a QZ-quality basis set from ref 75. [h]Gas-phase UV absorption spectra from ref 76. [i]From an absorption spectra in solid 3-methylpentane at 77 K (ref 77), unless noted otherwise. [j]From the Cotton−Mouton effect in a carbon tetrachloride solution at 293 K (ref 78). [k]The 0−0 transition from a gas-phase UV absorption spectrum (ref 79). [l]The 0−0 transition after extrapolation to the gas phase from fluorescence excitation spectra of matrix-isolated anthracene (ref 80). [m]Estimated reference value from ref 55.

change only by 1% (σ_D) and 3% (S_{CR}) if these additional data are included.

Table 2 shows mean signed errors and mean unsigned errors (MSE and MUE, respectively) obtained with the SMSSP model relative to reference data. When we use NPE(D) for bulk electrostatics, which we take as a standard procedure, the SMSSP error does not exceed 0.5 kcal/mol on average (independently of basis set); this is comparable to or even better than the error obtained using the SMD model with the same density functionals and basis sets. The SMSSP model that utilizes the NPE electrostatics based on *Gaussian 09*'s default choice of intrinsic Coulomb radii, i.e. the UFF radii[48] scaled by 1.1 [see the NPE(1.1 × UFF) entries in Table 2], gives a MUE of about 0.6 kcal/mol using the same σ_D and S_{CR} values as optimized for NPE(D).

Table 2 indicates that the SMSSP errors increase to ∼0.7 kcal/mol when one uses the GB bulk electrostatics instead of the NPE one. Reoptimization of σ_D and S_{CR} for the GB/mPW1PW/MG3S bulk electrostatics based on ChElPG charges and SM12 intrinsic Coulomb radii yields MUE = 0.56 kcal/mol and the new values of σ_D = −1.860 kcal mol^{-1} Å^{-3} and S_{CR} = 0.313, which are very close to the original SMSSP values (σ_D = −1.850, S_{CR} = 0.369). This reoptimized model may be called SMSSP(GB).

Note that the SMSSP model uses only two model parameters to account for non-bulk-electrostatic contributions over a set of 643 solvation data for solutes in solvents without hydrogen bonding whereas the SMD model[13] uses twelve for the same test set, and the SM12 model[14] uses 22 so-called atomic surface tension coefficients to describe such contributions (see, for

dx.doi.org/10.1021/ct400329u | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 5. Gas−Cyclohexane Solvatochromic Shifts by Various Methods ($\Delta\omega$, in cm^{-1})

		this work			
molecule	type	$\Delta\omega^a$	$\Delta\omega^b$	$\Delta\omega$ (RZ)[c]	$\Delta\omega$ (exptl.)
benzene	L_b	127	106	316	209,[d] 308[72]
	L_a	542	842	598	1070[72]
azulene	L_b	−123	−136	−162	−105[e]
	L_a	234	270	288	340[e]
naphthalene	L_b	307	326	332	275,[d] 389[e]
	L_a	1344	1707	879	367,[e] 902[d]
anthracene	L_b	575	575		
	L_a	2166	2471		866,[d] 1030[79]
coronene	L_b	870	870		209[79]
	L_a	1971	1971		800[79]

[a]The M06-2X and TDM06-2X values of $\Delta\omega_D$ from Table 4. [b]Averaged over all $\Delta\omega_D$ values given in Table 4 for each transition. [c]Computed by Rösch and Zerner (RZ)[82] and given here using the sign convention for solvatochromic shifts adopted in the present work. [d]In *n*-pentane (ref 81). [e]Averaged over 21 solvents with n^2 in the range of 1.89−2.27 using the data from ref 76.

example, Table 2 in ref 14 with the column containing the $\tilde{\sigma}_i^{[n]}$-type coefficients for the SM12 set of model parameters to be used in combination with ChElPG charges). One advantage of basing the result on polarizabilities is that the model should be more reliable when used outside the range of the original training set, since it does not require new parameters for new atomic numbers or new geometry dependences of the surface tensions to adequately treat new functional groups.

Figure 1 shows the difference in the magnitude of non-bulk-electrostatic effects predicted by the SMSSP model ($G_D + G_{CR}$) and by the SMD model (G_{CDS}). The SMSSP model predicts such contributions to be more negative than the SMD model does. Nevertheless, the absolute deviation is only about 13% of the *total* free energy of solvation, which is −4.5 kcal/mol as averaged over 643 reference data (in which the non-bulk-electrostatic effects dominate). The non-bulk-electrostatic contribution to the SMSSP free energy of solvation is −2.8 kcal/mol on average (−5.0 kcal/mol comes from dispersion and 2.2 kcal/mol from cavitation−repulsion) in comparison to the corresponding CDS contribution of the SMD model averaged over 643 data (−2.6 kcal/mol).

If we assume that all the solvation energies on the test set of 643 data are zero ($\Delta G_{EP} = G_D = G_{CR} = 0$), the mean unsigned error will be 4.5 kcal/mol. In the case when only bulk electrostatics are included ($G_D = G_{CR} = 0$), the MUE reduces to only 2.8 kcal/mol [if we use, for example, the NPE(D)/B3LYP/MG3S values of ΔG_{EP}], indicating that non-bulk-electrostatic effects are significant for the given solvation data. Including the solute−solvent dispersion term estimated by eq 4 without other contributions (i.e., $\Delta G_{EP} = G_{CR} = 0$) allows one to substantially reduce the error (MUE = 0.6 kcal/mol) by varying the value of only σ_D, indicating the significance of solute−solvent dispersion. The optimized value of σ_D in this case is −1.643 kcal mol^{-1} $Å^{-3}$ which is only by ∼12% smaller than the optimized SMSSP value (σ_D = −1.850 in the same units).

We have tested a model that defines the dispersion contribution G_D as a function of distributed atomic polarizabilities calculated using an algorithm introduced earlier[56] rather than a function of the total molecular polarizability. In this case, G_D is treated by the following formula:

$$G_D = \frac{n_s^2 - 1}{n_s^2 + 2} \sum_i \sigma_D^{(Z_i)} \alpha_M^{(i)} \tag{9}$$

where the index *i* runs over all atoms in the solute molecule, Z_i is the atomic number of atom *i*, and $\sigma_D^{(Z_i)}$ is an element-dependent empirical constant for the elements H, C, N, O, F, Si, S, P, Cl, Br, and I. The quantity $\alpha_M^{(i)}$ is the spherically averaged atomic polarizability of atom *i* in molecule *M*. By optimizing α_D^H, α_D^C, α_D^N, and α_D^O for the case of $\Delta G_{EP} = G_{CR} = 0$ and fixing the $\sigma_D^{(Z_i)}$ values for F, Si, S, P, Cl, Br, and I at −1.643 kcal mol^{-1} $Å^{-3}$ (see the previous paragraph), we obtain σ_D^H, σ_D^C, σ_D^N, and σ_D^O equal to −1.286, −2.020, −3.053, and −4.019 kcal mol^{-1} $Å^{-3}$, respectively, and we obtain MSE = 0.02 and MUE = 0.48 kcal/mol, which are somewhat smaller than in the case of a model using only one σ_D parameter for all elements and assuming that $\Delta G_{EP} = G_{CR} = 0$ for simplicity (in the latter case, MUE = 0.60 kcal/mol). Nevertheless, increasing a number of parameters from one to four increases the accuracy of the model only by 20%. We conclude that the use of eq 4 (with one dispersion parameter) is an adequate approximation.

Next, we compare the present results to PCM models in *Gaussian 09*. By default *Gaussian 09* does not add any non-bulk-electrostatic contribution to the NPE(1.1 × UFF) solvation energy; this leads to MUE ∼ 3 kcal/mol over the 643 solvation data used here. The NPE(1.1×UFF) bulk electrostatic energies are less negative on average than their NPE(D) counterparts. For example, the NPE(1.1×UFF)/B3LYP/MG3S bulk electrostatic energy averaged over 643 data equals −1.2 kcal/mol, whereas the corresponding NPE(D) energy is −1.8 kcal/mol.

Table 2 also compares the SMSSP errors to those obtained by the standard PCM treatment of cavitation, dispersion, and repulsion based on PC and FT calculations using *Gaussian 09*'s default settings (see section 3.3 for more details). The PCM cavitation, dispersion, and repulsion energies were obtained for 240 solute data in CCl_4, cyclohexane, and heptane, and they were combined with the NPE(1.1×UFF) and NPE(D) bulk electrostatic solvation energies calculated using the B3LYP/MG3S method. We have found that the standard PCM cavitation calculation in *Gaussian 09* overestimates the cavitation energy contribution, resulting in unphysically positive free energies of solvation. For example, the standard PCM calculation gives 14.8 kcal/mol on average for cavitation, and it gives −11.2 and 0.7 kcal/mol on average for dispersion and repulsion, respectively, in comparison with the *total* reference solvation energy averaged over 240 data (−5.0 kcal/mol). The bulk electrostatic contribution averaged over 240 data is equal to −1.4 kcal/mol with NPE(1.1 × UFF) and −2.0 kcal/mol with NPE(D).

In the case of PCMopt, the averaged cavitation, dispersion, and repulsion contributions are equal to 1.9, −5.0, and ∼0 kcal/mol, respectively. For comparison, the values of G_D and G_{CR} obtained using SMSSP and averaged over 240 data are equal to −5.2 and 2.2 kcal/mol, respectively. Thus, the PCMopt results are in better agreement with the SMSSP results than those obtained from the standard PCM treatment of cavitation, dispersion, and repulsion based on PC and FT calculations using *Gaussian 09*'s default settings. In addition, the use of the PCMopt method results in more accurate solvation free energies than in the case of the standard PCM calculations (see Table 2).

The reason for large cavitation energies obtained in the standard PCM calculation of cavitation energies is that such a calculation uses much larger values for the solvent radius (the

dx.doi.org/10.1021/ct400329u | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

solvent radius is set by default to 2.685−3.125 Å depending on the solvent) than the PCMopt calculation where the solvent radius is set to 1.3 Å for all solvents. The large cavitation energies used in standard PCM calculations contribute to the large positive mean signed errors previously pointed out[57] for PCM calculations on free energies of solvation of neutral molecules in organic solvents. The solute cavities used in dispersion−repulsion calculations do not depend on the solvent radius (the surface areas are given in the SI).

Figure 2 shows a satisfactory correlation between the values of G_D calculated using SMSSP (i.e., by eq 4) and those obtained using the FT approach[15,16] for dispersion and repulsion based on the PCMopt set of atomic radii of the solute atoms, although this is less of a range in the computed PCMopt FT values than in the SMSSP values.

Table 3 compares dispersion energies of various models for three straight-chain alkanes, in particular, for *n*-propane (C_3H_8), *n*-pentane (C_5H_{12}), and *n*-octane (C_8H_{18}). The solute−solvent dispersion energies for *n*-alkanes in water obtained by Floris and Tomasi[15] can be reproduced reasonably well by the standard PCM calculation of dispersion and repulsion using *Gaussian 09*'s default values for the van der Waals atomic radii of the solute atoms (i.e., by using the UFF radii[48]). The SMSSP dispersion energies are systematically less negative than the corresponding energies obtained in ref 15. The SMSSP dispersion energies are in better agreement with the reference solvation free energies[36] than the remaining entries in Table 3 when they are compared to experimental results directly based on the assumption that solute−solvent dispersion is the only contributor to the free energy of solvation in these cases or the remaining contributions (other than solute−solvent dispersion) are mutually canceled out. This is true even without making such an assumption when one amends the solute−solvent dispersion energies (Table 3) with the corresponding SMSSP estimates of G_{CR} and ΔG_{EP}. For example, the SMSSP cavitation−repulsion energy of *n*-C_8H_{18} in cyclohexane and water, G_{CR}, is 3.2 and 9.3 kcal/mol, respectively; the corresponding bulk electrostatic contribution, ΔG_{EP}, is −0.5 kcal/mol in cyclohexane and −1.3 kcal/mol in water (according to the SMD/B3LYP/MG3S calculation).

4.2. Excited States. Table 4 shows the gas-phase ground-state (GS) spherically averaged molecular polarizabilities (α_{GS}) of benzene, azulene, naphthalene, anthracene, and coronene and the excited-state (ES) polarizabilities (α_{ES}) for the lowest L_a and L_b $\pi\rightarrow\pi^*$ electronic states (using Platt's notations[58]) in the same molecules, along with the corresponding vertical excitation energies. The excited-state polarizabilities are given relative to the α_{GS} values. Ground-state calculations were performed using the density functionals M06[45,46] and M06-2X,[45,46] and the coupled cluster method with single and double excitations[59,60] (CCSD). The latter was employed using the frozen core approximation. Excited-state calculations were performed using time-dependent density functional theory[61−64] (TDDFT) and the equation-of-motion CCSD method[65,66] (EOM-CCSD). The MG3 basis set[47,54] was employed. The ground-state polarizabilities from DFT calculations were calculated analytically. The excited-state polarizabilities from TDDFT calculations were computed by numerical differentiation of analytical dipole moments in the external electric field (by using the option *Polar=Numer* in *Gaussian 09*), and the CCSD and EOM-CCSD polarizabilities were calculated by second-order numerical differentiation of energies (by using the option *Polar=DoubleNumer*). The reference values (both theoretical and experimental) for ground- and excited-state polarizabilities and vertical excitation energies are given when available.[55,67−80]

The gas−cyclohexane solvatochromic shifts given in Table 4 ($\Delta\omega_D$) were estimated from the gas-phase polarizabilities using the following equation:

$$\Delta\omega_D = -\sigma_D(\alpha^{ES} - \alpha^{GS})\frac{n^2 - 1}{n^2 + 2} \quad (10)$$

where α is a spherically averaged polarizability of the solute molecule in the ground or excited electronic state, n is the solvent's refractive index (n = 1.427), and σ_D is the SMSSP empirical parameter equal to −1.850 kcal mol^{-1} Å^{-3}. Note that the total solvatochromic shift (that includes dispersion and remaining contributions) is defined here as $\Delta\omega = \omega_{gas} - \omega_{sol}$ where ω_{gas} and ω_{sol} are excitation energies in the gas phase and in solution, respectively.

Assuming that the corresponding solvatochromic shifts due to bulk electrostatic effects for benzene, azulene, naphthalene, anthracene, and coronene will be small, we can compare the predicted dispersion shifts directly to experimental shifts[72,76,79,81] available for given solutes in cyclohexane or other solvents with similar values of ε and n. The SMSSP predicted and observed solvatochromic shifts are given in Table 5. Table 5 also shows solvatochromic shifts calculated by Rösch and Zerner using a solute−solvent dispersion model based on a configuration interaction treatment of solute and solvent molecules.[82]

Comparison of the results to experimental results is complicated because the experimental results also contain contributions from bulk electrostatics. However, qualitative agreement between observed solvatochromic shifts or those computed by Rösch and Zerner[82] and those predicted by SMSSP is generally good. In particular, the SMSSP model correctly predicts the larger red shift due to dispersion for the L_a transition in all studied compounds in comparison with the corresponding L_b transition. Our model correctly predicts the negative (blue) shift for the L_b transition in azulene. The solvatochromic shifts due to dispersion are usually red,[82] and this can be explained in the context of the SMSSP model by the excited-state polarizabilities usually being larger than the ground-state ones (i.e., $\alpha_{ES} > \alpha_{GS}$). In the case of the L_b transition in azulene, we observe the opposite situation ($\alpha_{ES} < \alpha_{GS}$).

5. SUMMARY

We present a model of the solute−solvent dispersion contribution to the free energy of solvation in the ground and excited states of solute molecules using a new solvation model, called SMSSP, based on state-specific polarizability. In addition to the solvent dielectric constant used in modeling electrostatic polarization, this model uses only three descriptors: the spherically averaged dipole polarizability of the solute molecule, the refractive index of the solvent, and the solvent's macroscopic surface tension. The formulation allows for description of dispersion−repulsion contributions to the solvation free energy using only two empirical parameters (in addition to atomic radii taken from previous work): one for repulsion and one for dispersion. We also tested the use of distributed models of the solute polarizability but found that the significantly increased number of parameters does not bring a significant increase in the accuracy of the method. This opens

the way to model dispersion and repulsion for any solute for which the polarizability is known or can be calculated. The SMD[13] and SM12[14] models are also defined for the whole periodic table, but earlier SM*x* models are defined only for selected atomic numbers. The SMSSP model is the first SM*x* model that is defined for excited electronic states as well as ground states.

Although the model was parametrized for non-hydrogen-bonding solutes, the parameters should also be applicable to predict the dispersion contribution when hydrogen bonding is present.

APPENDIX

In this appendix, we describe the functional form of eq 4.

By considering solvent molecules as spheres of radius R_S and a solute molecule as a sphere of radius R_M, we can use the London formula (LF)[8] for the dispersion interaction between two systems of spherical symmetry to model G_D as follows

$$G_D^{LF} = -\frac{3}{2}K_S\frac{I_M I_S}{I_M + I_S}\frac{\alpha_M\alpha_S}{(R_M + R_S)^6} \quad \text{(A1)}$$

where K_S is an effective number of solvent molecules in the first solvation shell, I_M and I_S are the first ionization potential of the solute M and the solvent S, respectively, and α_S is the spherically averaged molecular polarizability of the solvent molecule S. The latter polarizability may be derived using the Lorentz−Lorenz equation that relates the refractive index of a substance to its mean polarizability as follows

$$\alpha_S = \frac{n_S^2 - 1}{n_S^2 + 2}\frac{3W_S A_0}{4\pi\rho_S N_{Av}} \quad \text{(A2)}$$

where n_S is the refractive index of S in dimensionless units, ρ_S is the solvent's density in kg/m^3 (at 298 K), W_S is the solvent's molecular weight in kg/mol, N_{Av} is the Avogadro constant in mol^{-1}, and A_0 is the constant equal to 10^{30} if α_S is expressed in Å^3. The effective number of solvent molecules in the first solvation shell can be expressed as

$$K_S = \eta_S\frac{V_{MS} - V_M}{V_S} = \eta_S\frac{(R_M + 2R_S)^3 - R_M^3}{R_S^3} \quad \text{(A3)}$$

where η_S is the packing density of solvent molecules (one can assume a cubic packing yielding $\eta_S = (4/3)\pi R_S^3(8R_S^3)^{-1} = \pi/6$ or ≈ 0.5), V_{MS} is the volume of a supersolute including a bare solute molecule and all the solvent molecules in the first solvation shell, and V_M and V_S are the molecular volume of M and S, respectively. The solvent's radius can be obtained from its molecular weight and density as

$$R_S = \frac{1}{2}\left(\frac{W_S}{\rho_S N_{Av}}\right)^{1/3} \quad \text{(A4)}$$

The solute's radius can be obtained from the solute molecular surface area (S_M) calculated by the analytic surface area (ASA) algorithm using Bondi's van der Waals radii (see references in the text), as follows

$$R_M = \left(\frac{S_M}{4\pi}\right)^{1/2} \quad \text{(A5)}$$

Using eqs A2 and A3, eq A1 can be rewritten as

$$G_D^{LF} = \sigma_D^{LF}\alpha_M\frac{n_s^2 - 1}{n_s^2 + 2} \quad \text{(A6)}$$

where σ_D^{LF} is expressed as

$$\sigma_D^{LF} = -\eta_S\frac{(R_M + 2R_S)^3 - R_M^3}{R_S^3(R_M + R_S)^6}\frac{9W_S A_0}{8\pi\rho_S N_{Av}}\frac{I_M I_S}{I_M + I_S} \quad \text{(A7)}$$

The quantities used on the right-hand side of eq A7 are not used by the SMSSP model explicitly but they are included in the SMSSP empirical parameter σ_D (eq 4) *implicitly* through the SMSSP parametrization against experimental reference data.

One further issue to consider is whether it is appropriate to use the entire molecular polarizability in eq 4 since any portion of the polarizability due to buried atoms will not contribute strongly to dispersion interactions with the solvent. However, we have found[56] that almost all of the molecular polarizability is located on surface of the molecule, with hardly any in the interior.

ASSOCIATED CONTENT

Supporting Information

Reference solvation free energies and computational data. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: cramer@umn.edu (C.J.C) and truhlar@umn.edu (D.G.T.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported in part by the U.S. Army Research Laboratory under grant no. W911NF09-100377 and by the U.S. Department of Energy, Office of Basic Energy Sciences, under SciDAC grant no. DE-SC0008666.

REFERENCES

(1) Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertrán, J., Eds.; Kluwer: Dordrecht, The Netherlands, 1996; In *Understanding Chemical Reactivity*; Springer: New York, Vol. *17*, pp 1−80.

(2) Curutchet, C.; Cramer, C. J.; Truhlar, D. G.; Ruiz-López, M. F.; Rinaldi, D.; Orozco, M.; Luque, F. G. *J. Comput. Chem.* **2003**, *24*, 284−297.

(3) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760−768.

(4) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2009**, *42*, 493−497.

(5) Jeziorski, B. J.; Moszynski, R.; Ratkiewicz, A.; Rybak, S.; Szalewicz, K.; Williams, H. L. In *Methods and Techniques in Computational Chemistry: METECC-94*; Clementi, E., Ed.; STEF: Cagliari, Italy, 1993; Vol. *B*, p 79.

(6) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.

(7) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3094.

(8) Eisenschitz, R.; London, F. *Z. Phys.* **1930**, *60*, 491−527.

(9) Koide, A.; Meath, W. J.; Allnatt, A. R. *Chem. Phys.* **1981**, *58*, 105−119.

(10) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726−3741.

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

(12) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305−8311.

(13) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.
(14) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 609−620.
(15) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616−627.
(16) Floris, F.; Tomasi, J.; Pascual-Ahuir, J. L. *J. Comput. Chem.* **1991**, *12*, 784−791.
(17) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.
(18) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision C.01; Gaussian, Inc.: Wallingford, CT, 2009.
(19) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717−726.
(20) Claverie, P. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Veillard, A., Eds.; Reidel: Dordrecht, The Netherlands, 1982; Vol. 3, pp 151−175.
(21) Linder, B. In *Intermolecular Forces*, Hirschfelder, J. O., Ed.; John Wiley & Sons, Inc.: New York, 1967; In *Advances in Chemical Physics*; Springer: New York; Vol. *12*, pp 225−282.
(22) Aguilar, M. A.; Olivares del Valle, F. J. *Chem. Phys.* **1989**, *138*, 327−336.
(23) Rauhut, G.; Clark, T.; Steinke, T. *J. Am. Chem. Soc.* **1993**, *115*, 9174−9181.
(24) Tuñón, I.; Silla, E.; Pascual-Ahuir, J. L. *Chem. Phys. Lett.* **1993**, *203*, 289−294.
(25) Amovilli, C.; Mennucci, B. *J. Phys. Chem. B* **1997**, *101*, 1051−1057.
(26) Shimizu, K.; Freitas, A. A.; Farah, J. P. S.; Dias, L. G. *J. Phys. Chem. A* **2005**, *109*, 11322−11327.
(27) Curutchet, C.; Orozco, M.; Luque, F. J.; Mennucci, B.; Tomasi, J. *J. Comput. Chem.* **2006**, *27*, 1769−1780.
(28) Basilevsky, M. V.; Leontyev, I. V.; Luschekina, S. V.; Kondakova, O. A.; Sulimov, V. B. *J. Comput. Chem.* **2006**, *27*, 552−570.
(29) Cossi, M.; Barone, V. *J. Chem. Phys.* **2000**, *112*, 2427−2435.
(30) Weijo, V.; Mennucci, B.; Frediani, L. *J. Chem. Theory Comput.* **2010**, *6*, 3358−3364.
(31) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117−9133.
(32) Liotard, D. A; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1995**, *16*, 422−440.
(33) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441−451.
(34) Mantina, M.; Chamberlin, A. C.; Valero, R.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2009**, *113*, 5806−5812.
(35) Mantina, M.; Valero, R.; Cramer, C. J.; Truhlar, D. G. In *CRC Handbook of Chemistry and Physics*, 91st ed. (2010−2011); Haynes, W. M., Ed.; CRC Press: Boca Raton, FL, 2010; pp 9-49−9-50.
(36) Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. *Minnesota Solvation Database version 2012*; University of Minnesota: Minneapolis, MN, 2012. http://comp.chem.umn.edu/mnsol (accessed November 26, 2012).
(37) Winget, P.; Dolney, D. M.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *Minnesota Solvent Descriptor Database version 1999*; University of Minnesota: Minneapolis, MN, 1999. http://comp.chem.umn.edu/solvation/mnsddb.pdf (accessed September 21, 2012).
(38) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032−3041.
(39) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3110.
(40) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(41) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(42) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(43) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664−675.
(44) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101/1−18.
(45) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(46) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157−167.
(47) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384−1388. See also http://comp.chem.umn.edu/basissets/basis.cgi (accessed July 16, 2012).
(48) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(49) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* 1989, *157*, 164−170 and references therein.
(50) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(51) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.
(52) Olson, R. M.; Marenich, A. V.; Chamberlin, A. C.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Winget, P.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM*, version 2012; University of Minnesota: Minneapolis, MN, 2012.
(53) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
(54) Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. *Chem. Phys. Lett.* **1999**, *306*, 407−410.
(55) Miller, T. M. Atomic and Molecular Polarizabilities. In *CRC Handbook of Chemistry and Physics*, 93rd ed. (2012−2013); Haynes, W. M., Ed.; CRC Press: Boca Raton, FL, 2012; p 10−186; online version at http://www.hbcpnetbase.com (accessed December 7, 2012).
(56) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *Chem. Sci.* **2013**, *4*, 2349−2356.
(57) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2011−2033.
(58) Platt, J. R. *J. Chem. Phys.* **1949**, *17*, 484−495.
(59) Bartlett, R. J.; Purvis, G. D. *Int. J. Quantum Chem.* **1978**, *14*, 561−581.
(60) Scuseria, G. E.; Janssen, C. L.; Schaefer, H. F., III. *J. Chem. Phys.* **1988**, *89*, 7382−7387.
(61) Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, *52*, 997−1000.
(62) Casida, M. E. In *Time-Dependent Density-Functional Response Theory for Molecules*; Chong, D. P., Ed.; World Scientific: Singapore, 1995; Vol. *1*, p 155.
(63) Bauernschmitt, R.; Ahlrichs, R. *Chem. Phys. Lett.* **1996**, *256*, 454−464.
(64) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1998**, *109*, 8218−8224.

(65) Koch, H.; Jørgensen, P. *J. Chem. Phys.* **1990**, *93*, 3333−3344.
(66) Stanton, J. F.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 7029−7039.
(67) Norman, P.; Jonsson, D.; Ågren, H. *Chem. Phys. Lett.* **1997**, *268*, 337−344.
(68) Christiansen, O.; Hättig, C.; Jørgensen, P. *Spectrochim. Acta, Part A* **1999**, *55*, 509−524.
(69) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W. *J. Chem. Phys.* **2008**, *128*, 134110/25.
(70) Heitz, S.; Weidauer, D.; Rosenow, B.; Hese, A. *J. Chem. Phys.* **1992**, *96*, 976−981.
(71) Okruss, M.; Müller, R.; Hese, A. *J. Chem. Phys.* **1999**, *110*, 10393−10402.
(72) Bayliss, N. S.; Hulme, L. *Aust. J. Chem.* **1953**, *6*, 257−277.
(73) Pantos, E.; Philis, J.; Bolovinos, A. *J. Mol. Spectrosc.* **1978**, *72*, 36−43.
(74) Hiraya, A.; Shobatake, K. *J. Chem. Phys.* **1991**, *94*, 7700−7706.
(75) Chong, D. P. *Can. J. Chem.* **2010**, *88*, 787−796.
(76) Weigang, O. E., Jr.; Wild, D. D. *J. Chem. Phys.* **1962**, *37*, 1180−1187.
(77) Mathies, R.; Albrecht, A. C. *J. Chem. Phys.* **1974**, *60*, 2500−2508.
(78) Cheng, C. L.; Murthy, D. S. N.; Ritchie, G. L. D. *Aus. J. Chem.* **1972**, *25*, 1301−1305.
(79) Morales, R. G. E. *J. Phys. Chem.* **1982**, *86*, 2550−2552.
(80) Wolf, J.; Hohlneicher, G. *Chem. Phys.* **1994**, *181*, 185−208.
(81) Robertson, W. W.; Weigang, O. E., Jr.; Matsen, F. A. *J. Mol. Spectrosc.* **1957**, *1*, 1−10.
(82) Rösch, N.; Zerner, M. C. *J. Phys. Chem.* **1994**, *98*, 5817−5823.