JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Field-Extremum Model for Short-Range Contributions to Hydration Free Energy

Anna Pomogaeva and Daniel M. Chipman*

Radiation Laboratory, University of Notre Dame, Notre Dame, Indiana 46556-5674, United States

Supporting Information

ABSTRACT: The performance in describing hydration free energies of a broad class of neutral, cationic, and anionic solutes is tested for the recently proposed FESR (Field-Extremum Short-Range) implicit solvation model for interactions between the solute and nearby water molecules, as taken in conjunction with the previously developed SS(V)PE (Surface and Simulation of Volume Polarization for Electrostatics) dielectric continuum model for long-range interactions with bulk water. The empirical FESR model mainly describes solute−water hydrogen bonding interactions by correlating them with the maximum and minimum values of the electric field produced by the solute at the surface of the cavity that excludes solvent. A preliminary report showed that, with only four adjustable parameters, the FESR model, in conjunction with SS(V)PE, can produce hydration energies comparable to the best analogous efforts in the literature that utilized many more parameters. Here, the performance of the FESR model is more fully documented in several respects. The dependence on the underlying quantum mechanical method used to treat the internal electronic structure of the solute is tested by comparing uncorrelated Hartree−Fock to correlated density functional calculations and by comparing a modest sized to a large basis set. The influence of cavity size is studied in connection with an isodensity contour construction of the cavity. The sensitivity of the results to the parameters in the FESR model is considered, and it is found that the dependence on the electric field strength is quite nonlinear, with an optimum exponent consistently in the range of 3 to 4. Overall, it is concluded that the FESR model shows considerable utility for improving the accuracy of implicit models of aqueous solvation.

INTRODUCTION

In the gas phase, it is now possible to accurately calculate properties and chemical reactions in many molecular systems using high level quantum mechanical (QM) methods. However, in condensed phases the environment impacts most chemical features so that such calculations in solution are still challenging and generally require a compromise between the theoretical reliability of an approach and its practical usefulness. Consequently, a wide variety of approaches has been developed to treat the effects of solvation. Direct attacks that explicitly treat many solvent molecules are capable of high accuracy but are computationally expensive. They remain expensive even if the solvent molecules are treated very approximately, as with molecular mechanical methods, due both to the necessity of including many solvent molecules to properly incorporate the long-range electrostatic solute−solvent interactions and to the need to carry out extensive statistical averaging over thermal motions in the solvent. Therefore, considerable interest remains in the development and improvement of highly simplified models that implicitly incorporate the most important influences of solvent on a solute. Several excellent reviews[1−3] are available that comprehensively discuss implicit solvation models.

One particularly important property is the free energy of solvation, which governs solubility, thermodynamic stability, pK_a, redox potential, and many other chemical attributes of a solute. For polar solutes in polar solvents, the long-range electrostatic interactions between solvent and solute dominate the solvation energy, so it is first necessary to have as good a description of these interactions as possible. For this purpose, most implicit solvation approaches use a classical dielectric continuum treatment based either on Poisson's equation[4] or on a generalized Born approach[5] to describe the statistically averaged response of bulk solvent.

Even with an accurate accounting of the long-range electrostatic interactions between solute and solvent, there generally remain significant short-range interactions that should be treated separately.[6−21] These short-range interactions are usually phenomenologically divided into nonelectrostatic dispersion and exchange-repulsion energies and also the differences between the short-range electrostatic interactions that actually occur with nearby solvent molecules from those that are described by the dielectric continuum model. Additional short-range interactions such as cavitation work, charge transfer effects, three-body interactions, and nonlinear solvent response are also sometimes explicitly distinguished.[1−3] Hydrogen bonding, while technically mostly arising from a combination of the aforementioned effects with perhaps also contributions from partial covalent character,[22−30] is sufficiently unique that it is often considered a separate interaction.

In a preliminary report,[31] we have described a very promising simple model having only a handful of parameters that implicitly describe some of the most important short-range effects of aqueous solvation. Used in conjunction with a modern dielectric continuum model of the long-range electrostatic effects, it provides quite reasonable hydration energies for a large collection of neutral and ionic solutes. In the present work, we further document this new model to show how it performs under different conditions of the calculation.

Received: August 16, 2011
Published: October 14, 2011

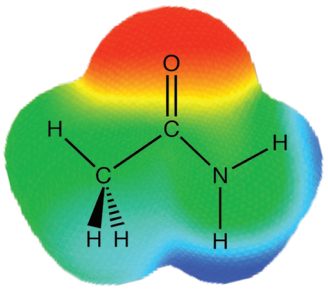

Figure 1. Acetamide molecule, colored according to the outgoing normal electric field F it generates on its $\rho_0 = 0.001$ au isodensity surface. The hottest red spot above the carbonyl oxygen is the position of F_{min}, where acetamide is likely to accept a hydrogen bond from a solvent water molecule. The hottest blue spot below one of the amide hydrogens is the position of F_{max}, where acetamide is likely to donate a hydrogen bond to a solvent water molecule.

Our preliminary report introduced the empirical FESR (Field Extremum Short-Range) model[31] that was argued to mainly describe hydrogen-bonding effects on hydration energies. Earlier studies[32−34] had indicated that the strength of a hydrogen bond depends on the electric field in its vicinity. In line with this, the FESR method basically searches the electric field produced by the solute at the cavity surface to find the hottest spots where hydrogen bonds can be expected to occur. Thus, an anion should accept a hydrogen bond from water at the position of minimum (i.e., most negative) outgoing normal electric field on the surface, while a cation formed by protonation of a neutral should donate a hydrogen bond to water at the position of maximum outgoing normal electric field on the surface. Neutral solutes should show similar although probably diminished effects. Furthermore, the strength of these hydrogen bonds should be related to the magnitude of the field at those positions. These ideas are illustrated in Figure 1.

The FESR model quantifies these ideas and provides a simple means with few parameters to describe short-range contributions to hydration energies. Together with a modest-level QM treatment of the solute and dielectric continuum treatment of long-range electrostatic contributions, the FESR model with only four adjustable parameters showed[31] mean unsigned errors (MUE) from experiment of only 0.9 kcal/mol for 264 neutral solutes and 2.4 kcal/mol for 111 ionic solutes, which are comparable to the best analogous previous efforts in the literature with SM6 and SMVLE models that utilize many more parameters.[21]

The simplicity and accuracy of the FESR model makes it attractive for practical calculations, but the empirical nature of the approach demands proofs of its general reliability. In the present contribution, we explore the performance of the FESR model in several respects, including the influence of electron correlation and basis set size in the QM method used to obtain the internal electronic structure of the solute, the dependence on the size of the solute cavity, and the sensitivity to the adjustable parameters. Our development of this empirical model should also provide guidance into how a more fundamental implicit model of short-range solvation effects may be developed in future work.

METHOD AND COMPUTATIONAL DETAILS

Experimental Comparison. Testing the accuracy of the parametrization of any semiempirical model requires a pool of available experimental data. In the present research, experimental hydration free energies were taken from the Minnesota Solvation Database, version 2009,[35] that has been collected and kindly provided as an open source database by the Minnesota group. The standard state assumed there and here corresponds to the Ben-Naim convention[36] of identical concentrations in gas and in solution. Ionic hydration free energies are based on the Tissandier et al.[37] value of −265.9 kcal/mol for the proton hydration free energy. The database developers estimate uncertainties in the experimental data to be about 0.2 kcal/mol for neutrals and about 3 kcal/mol for ions.[35]

This database also includes optimized gas-phase geometries for all solutes. We adopted these geometries in our work as unchanged upon solvation. In principle, solvation energies will be affected by changes in geometry and zero-point vibration energy upon passing to the condensed phase,[38] but the solutes in this collection are not expected to undergo significant structural changes in solution. Therefore, it is believed that inclusion of such changes would not significantly affect our results at the level of accuracy being considered. Allowing the solutes to relax their geometries and change their vibrational frequencies in solution would presumably tend to further increase the accuracy of the results, although at the expense of sacrificing much of the simplicity of the approach.

This Minnesota Solvation Database contains 274 neutrals, 52 cations, and 60 anions, overall making 386 aqueous solutes. However, only 372 of them, including 261 neutrals, 51 cations, and 60 anions, were successfully treated by the single-center integration scheme used in our dielectric continuum program, which fails for severely nonspherical solutes. It should be noted that this is not a shortcoming of the model being presented but rather is only a limitation of the particular surface integration scheme implemented in the current version of the computer code that was utilized. The calculations failed either at the smallest cavity size or at one of the QM levels considered for 13 neutrals and one cation. Several of the solutes were successfully treated only after shifting the cavity surface integration center from the default origin at the center of nuclear charge to some other point. The full list of failed solutes and of those with a reset center of surface integration can be found for each computational case in Tables S1 and S2 of the Supporting Information. Note that our earlier report[31] that used only one QM method and cavity size successfully treated three more neutrals for a total of 375 solutes. While the extent of the database used influences the empirical fitting of parameters, no significant differences exist between the results from there and here.

Quantum Mechanical Methods. Several combinations of QM methods were used to treat the internal electronic structure of the solute.

The effects of electron correlation in the solute were investigated by comparing uncorrelated Hartree−Fock (HF) *ab initio* methods with correlated B3LYP[39−41] density functional methods.

The influence of the basis set was considered by comparing the modest-sized 6-31+G*[42−44] basis with the rather bigger G3large[45] basis set. The former is double-ζ in the valence space and also has polarization and diffuse functions on all atoms but hydrogen. The latter is of triple-ζ quality in the valence space, has multiple polarization functions on all atoms, has diffuse functions on all atoms, and is fortunately available for all of the atoms included in the database. We remark that additional diffuse functions are not deemed necessary, in part because it has been found that application of a dielectric continuum tends to make anions smaller.[46,47] In this connection, we note that application of a dielectric

continuum also tends to make cations larger,[47,48] both of these trends being counterintuitive.

Dielectric Continuum Calculations. Long-range electrostatic contributions to the hydration free energy were calculated by means of the SS(V)PE (Surface and Simulation of Volume Polarization for Electrostatics) dielectric continuum model.[49,50] This model pays particular attention to approximating the volume polarization arising from the long-range electrostatic effects of solute charge density that penetrates outside the cavity nominally enclosing it, the effects of which are treated more exactly in the SVPE (Surface and Volume Polarization for Electrostatics) model.[51−55] If implemented with the same cavity, the SS(V)PE method becomes equivalent[56,57] to the modified version of the IEF-PCM method that is often currently used.[58,59]

To link a QM treatment of the solute with a classical dielectric continuum model of solvent, it is important to choose a proper physical boundary between the solute and the solvent, i.e., the cavity which excludes solvent. Unlike many other implementations that use atom-centered spheres to define the solute cavity, we prefer to use a cavity based on an electronic isodensity contour of the solute. This has the advantage of being naturally adapted to the solute shape and of requiring only a single parameter needed to determine the overall cavity size. For this work, we examined three different cavity sizes, corresponding to solute electronic isodensity contours of ρ_0 = 0.0005, 0.001, and 0.002 au, which were found to be reasonable in previous studies.[60,61] Surface integrations over the cavity were carried out by single-center Lebedev quadrature with 1202 grid points. The water dielectric constant was set to 78.36.

Short-Range Contributions. After determination of $\Delta G_{SS(V)PE}$ for the dielectric continuum contributions to hydration free energies, an additional correction term for short-range contributions is calculated by means of the FESR method.[31]

$$\Delta G_{exptl} \approx \Delta G_{SS(V)PE} + \Delta G_{FESR} \quad (1)$$

Here, ΔG_{FESR} is an empirically parametrized term depending on the minimum and maximum values of the outgoing normal electric field produced by the solute anywhere on the cavity surface.

$$\Delta G_{FESR} = a + b|F_{min}|^d + cF_{max}^d \quad (2)$$

The value of F_{min} is taken as the lower of zero or the most negative value of the outgoing normal electric field, and F_{max} is the higher of zero or the most positive value of the outgoing normal electric field. The empirical parameters a, b, c, and d are adjusted to minimize the least-squares deviation from experimental values. Their values are reported here to yield free energy in kilocalories per mole when the electric fields are expressed in atomic units. Unless specified otherwise, equal weight is given to all solutes in the training set. In some particular cases, we also tested a fitting strategy that utilized different weights according to the estimated experimental error for each solute.

As with many other treatments of short-range effects,[62−64] the approach of eq 2 only includes the short-range interactions as a sum of independently calculated additive contributions to the solvation energy. Ideally, the short-range interactions should be allowed to polarize the electronic structure of the solute by including them as potential energy terms in the solute Hamiltonian,[65−67] in a manner similar to the way the long-range dielectric effects are usually included. One of our eventual goals is to adapt the FESR approach to produce such a potential energy term in the solute Hamiltonian in order to treat the influence of short-range solute−solvent interactions in a self-consistent reaction field framework.

The fitting of parameters in the FESR model was done for all 12 combinations from each of the three cavity sizes with ρ_0 = 0.0005, 0.001, and 0.002 au contours together with each of the four QM levels corresponding to HF/6-31+G*, B3LYP/6-31+G*, HF/G3large, and B3LYP/G3large.

All of the computations were made using a locally modified version of the GAMESS software.[68]

RESULTS AND DISCUSSION

General Remarks. For the most part, results will be discussed only for broad groups, such as the class of all solutes taken together or for the individual subclasses of neutrals, cations, and anions. For reference purposes, Tables S3−S6 of the Supporting Information give full details of the electric field extrema on the cavity surface, $\Delta G_{SS(V)PE}$, ΔG_{FESR}, and the related errors for each individual solute from the various combinations of computational methods considered.

The overall mean unsigned errors (MUE) for calculated hydration free energies compared to experiment are given in Table 1 for various combinations of QM methods and cavity sizes. Uncorrected SS(V)PE results are also given for reference purposes. The SS(V)PE results generally show the smallest errors for neutrals, somewhat larger errors for cations, and very large errors for anions. The SS(V)PE+FESR results are shown first for allowing the d parameter to optimize separately for each individual QM method and then for constraining the d parameter to have a common universal value for all of the QM methods. It is immediately evident that in all cases the FESR correction substantially reduces the errors over SS(V)PE alone, most particularly for anions.

The SS(V)PE+FESR MUE results for neutrals are generally somewhat larger than the estimated experimental errors of about 0.2 kcal/mol, the MUE results for cations are either comparable to or in some instances significantly lower than the estimated experimental errors of about 3 kcal/mol, and the MUE results for anions are either comparable to and in some instances slightly lower than the estimated experimental errors of about 3 kcal/mol. The most obvious explanation for how some MUE results can be lower than the estimated experimental errors is that the latter estimates are actually too high. Several other possible explanations of this matter have been previously elaborated.[31]

The results in Table 1 derive from giving equal weights to all solutes in the fitting process. Allowing different weights for different classes of solutes according to their estimated experimental uncertainties (i.e., relative weights of 1/0.2 for neutrals and 1/3 for ions) leads to quite similar results. For example, for B3LYP/6-31+G* with ρ_0 = 0.001 au, the SS(V)PE+FESR MUE from different weighting is 1.40 kcal/mol for all solutes, as compared to 1.37 kcal/mol obtained from equal weighting.

A convenient visualization of these results is given in Figure 2 for the particular case of B3LYP/6-31+G* calculations with ρ_0 = 0.001 au. All of the other various computational combinations considered also show quite similar behavior. The results in Figure 2 would all fall on the diagonal solid lines if perfect agreement with experimental results were achieved. The top left panel shows hydration free energies calculated from the SS(V)PE model alone compared to experimental results for all solutes, while the top right panel shows an expanded version for the neutral solutes. The agreement with experimental results is seen

Table 1. Mean Unsigned Errors in kcal/mol Obtained from Several QM Methods and Cavity Sizes[a]

solvation method	parameter d status[b]	all solutes (372)	neutrals (261)	cations (51)	anions (60)
		HF/6-31+G*			
SS(V)PE		4.65, 3.92, 5.28	1.65, 2.56, 4.63	7.13, 2.28, 6.82	15.61, 11.21, 6.77
SS(V)PE+FESR	individual d	1.73, 1.54, 2.03	1.19, 1.16, 1.67	3.24, 2.14, 3.11	2.80, 2.69, 2.66
SS(V)PE+FESR	universal d	1.76, 1.58, 2.03	1.26, 1.24, 1.69	3.05, 2.14, 3.07	2.83, 2.58, 2.64
		B3LYP/6-31+G*			
SS(V)PE		5.19, 3.98, 4.85	1.51, 1.81, 3.45	8.37, 2.95, 5.91	18.50, 14.32, 10.07
SS(V)PE+FESR	individual d	1.65, 1.37, 1.71	1.09, 0.94, 1.15	2.99, 1.80, 2.95	2.97, 2.88, 3.12
SS(V)PE+FESR	universal d	1.65, 1.37, 1.71	1.09, 0.94, 1.15	2.99, 1.80, 2.95	2.97, 2.88, 3.12
		HF/G3large			
SS(V)PE		4.86, 3.65, 4.65	1.63, 1.94, 3.68	8.00, 2.80, 6.43	16.22, 11.83, 7.33
SS(V)PE+FESR	individual d	1.93, 1.71, 2.15	1.52, 1.45, 1.83	2.99, 2.01, 3.29	2.81, 2.57, 2.61
SS(V)PE+FESR	universal d	1.93, 1.71, 2.15	1.52, 1.45, 1.84	2.99, 2.01, 3.26	2.81, 2.57, 2.60
		B3LYP/G3large			
SS(V)PE		5.43, 4.07, 4.52	1.61, 1.66, 3.01	9.26, 3.87, 5.20	18.80, 14.69, 10.50
SS(V)PE+FESR	individual d	1.71, 1.45, 1.66	1.15, 1.00, 1.11	3.01, 1.97, 2.69	3.04, 2.95, 3.17
SS(V)PE+FESR	universal d	1.73, 1.48, 1.67	1.14, 1.00, 1.09	3.18, 2.04, 2.74	3.05, 3.06, 3.27

[a] The three numbers in each slot correspond to results for cavity sizes from ρ_0 = 0.0005, 0.001, and 0.002 au, respectively. [b] "Individual d" means the parameter d is optimized individually for each QM method. "Universal d" means the parameter d is constrained to have a common universal value for all of the QM methods.

to be reasonable for neutrals and cations, while large and systematic errors are found for anions. The bottom panels show analogous results for the SS(V)PE+FESR model. Some notable improvement is found for neutrals and cations, and dramatic improvement is obtained for anions.

It is generally believed that solute surface area and/or volume are significant descriptors of generic short-range interactions between the solute and solvent.[69−77] However, our computational experiments (data not shown) found that further inclusion of terms proportional to solute surface area or volume led to only tiny improvements in the MUE. We surmise that the constant term a in the FESR model likely describes in some rough overall average sense such size-dependent contributions. It may then be that the solutes in the database used for training do not cover a sufficient range of sizes to specifically observe beyond this the influence of solute size over whatever other errors remain. In this connection, it may be noted that studies of cavitation have shown that in addition to size, solute shape is also a significant complicating factor.[78−80]

While it is believed that the field-dependent terms in the FESR model mainly describe hydrogen bonding effects, it should be pointed out that the empirical optimization of parameters necessarily also implicitly incorporates some contributions from dispersion and exchange-repulsion interactions in the model.

For some solutes, there are several hot spots where hydrogen bonds might be expected to occur. The FESR method takes care of such cases only insofar as the accumulated effects of all hydrogen bonds that are incorporated in the experimental values affect the values of the fitting parameters. A preliminary attempt to include more than one contribution to ΔG_{FESR} for some representative solutes having more than one local region of minimum and/or maximum electric field actually led to poorer agreement with experimental results. While this was done without a reoptimization of parameters, it is anyway clear that this matter is more subtle than it seems and will require further research to understand properly.

As a control experiment, we also tested the possibility that the FESR model might perform well alone, without any contributions from a dielectric continuum model. Even in the best case, this led to a very large MUE for neutrals of 9 kcal/mol and for ions of 20 kcal/mol, thus soundly refuting the possibility.

Dependence on QM Method. The QM method used to determine the electronic structure of the solute affects the hydration energy calculated in this work in two distinct ways. First, the QM method determines the solute size and shape through specifying the isodensity surface used to define the cavity. In this connection, we note that inclusion of electron correlation typically changes the cavity size by only about 0.01 Å (usually, but not always, increasing it), which is very small relative to the changes in cavity size of typically about 0.2 Å in passing from a ρ_0 of 0.0005 to one of 0.001 or from 0.001 to 0.002 au. Extension of the basis set has even less of an effect on the cavity size, usually slightly reducing it by less than 0.001 Å with HF and less than 0.004 Å with B3LYP. Therefore, any trends in results from different QM methods that are consistent at all cavity sizes examined are not due to the differences in how the various QM methods determine the cavity. Second, the QM method determines the electric properties presented by the solute at the cavity surface, specifically the electrostatic potential that is utilized in the SS(V)PE calculation and the electric field that is utilized in the FESR correction, and this is then the main route through which the QM method influences the results.

It is seen in Table 1 that each of the QM methods considered leads to good overall results. Even so, there are some notable small differences in performance that are attributable to electron correlation and basis set effects.

Inclusion of electron correlation, as done here by comparison of B3LYP to HF, should be expected to improve the results. With both basis sets, it is found that inclusion of electron correlation does indeed improve the SS(V)PE+FESR agreement with experimental results.

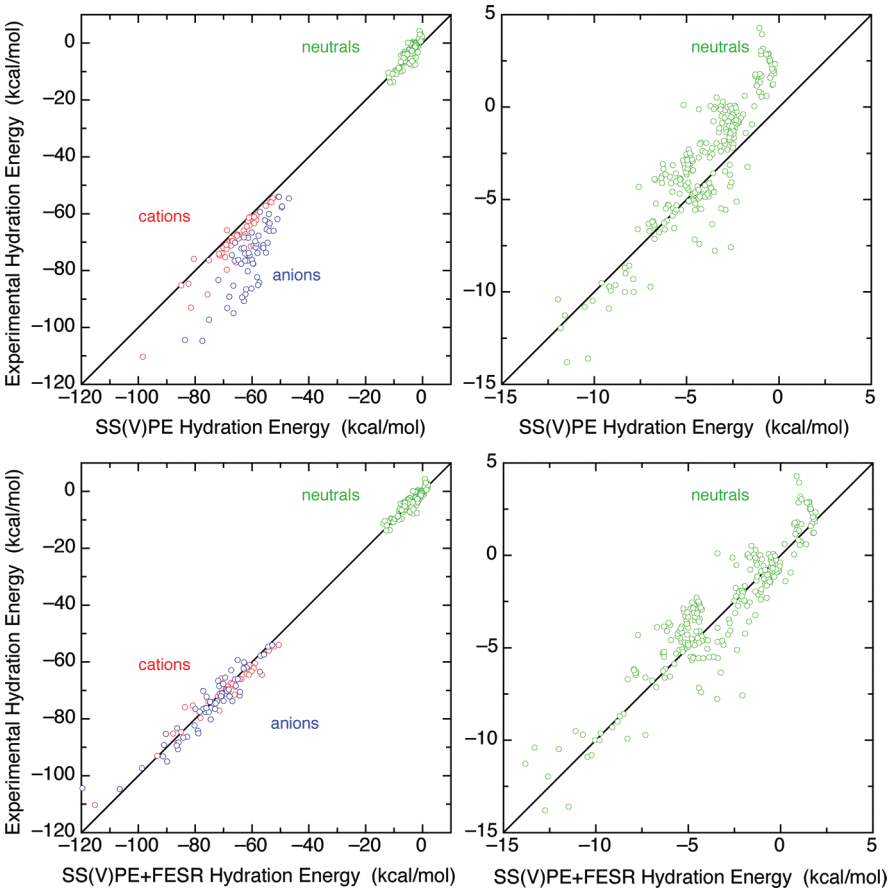

Figure 2. Comparison of experimental hydration energies with those calculated from the B3LYP/6-31+G* method with the $\rho_0 = 0.001$ au cavity size. The left panels show SS(V)PE and SS(V)PE+FESR results for all solutes, respectively, while the right panels show expanded versions covering only neutral solutes.

Extension of the basis set, as done here by comparison of G3large to 6-31+G*, should also be expected to improve the results. But with both HF and B3LYP, it is found that basis set extension instead slightly degrades the agreement with experimental results. A possible explanation for this apparently anomalous behavior of basis set extension is that the modest 6-31+G* basis set may accidentally tend to give better solute electric properties in the cavity surface region than does the larger G3large basis set. There is clearly a need for further research on this point to determine if other large basis sets give similar behavior.

Dependence on FESR Fitting Parameters. Examination of the optimum values of the parameters *a*, *b*, *c*, and *d* reported for all computational combinations in Table 2 shows that the value of *b* is always more than an order of magnitude larger than that of *c*. This is consistent with the substantially larger errors from SS(V)PE itself for anions as compared to cations or neutrals, which therefore requires larger FESR corrections for anions.

As expected on physical grounds discussed in the Introduction and seen from the detailed results in Tables S3−S6 of the Supporting Information, the FESR term involving F_{min} makes its largest contribution to anions; the term involving F_{max} makes its largest contribution to cations. Both contribute significantly to polar neutrals. In addition to the field-dependent terms, the FESR model also includes a constant term *a* which likely describes contributions from cavitation, exchange repulsion, and dispersion in some rough overall average sense.

Examples of the sensitivity of the results to *d* are shown in Figure 3, which correspond to the case of *d* being optimized separately for each individual QM method, while *a*, *b*, and *c* are reoptimized for each value of *d*. The top panel gives B3LYP/6-31+G* results for each cavity size as a function of *d*, showing that changes on the order of one unit in *d* lead to a significant change in the MUE. The bottom panel of Figure 3 gives results from different QM methods for the $\rho_0 = 0.001$ au cavity size as a function of *d* and again shows a similar sensivity to *d*.

Interestingly, experiments allowing F_{min} and F_{max} to have different values of *d* (data not shown) led to negligible improvement in the overall MUE. This implies that both cations and anions prefer the same value of *d*, suggesting it has some underlying physical significance.

It is seen in Table 2 that the optimum *d* varies only from 2.9 to 4.1 for all of the combinations of methods considered, and over an even smaller range for the various QM methods at each particular cavity size. This suggests that it might be provident to claim a universal value of *d* that is independent of the QM method, although still dependent on the cavity size and separate optimization for each QM method of the *a*, *b*, and *c* parameters. Table 1 includes the results from assuming that all of the QM methods share a common universal value of *d*, where it is seen that this constraint sacrifices less than (and usually much less than) 0.10 kcal/mol in MUE over allowing each QM method to have its own optimum *d*. The optimum parameter values under

Table 2. Optimum Parameter Values Obtained from Several QM Methods and Cavity Sizes, Allowing Each QM Method to Have Its Own Individual Optimum *d* at Each Cavity Size

parameter	ρ_0 = 0.0005 au	ρ_0 = 0.001 au	ρ_0 = 0.002 au
	HF/6-31+G*		
a	1.850	2.828	5.238
b	-0.378×10^6	-1.319×10^6	-0.491×10^6
c	-2.029×10^4	-1.897×10^4	0.066×10^4
d	3.4	4.1	4.0
	B3LYP/6-31+G*		
a	1.570	2.345	4.158
b	-0.259×10^6	-0.585×10^6	-0.523×10^6
c	-1.543×10^4	-1.107×10^4	0.027×10^4
d	3.1	3.6	3.8
	HF/G3large		
a	1.157	2.094	4.251
b	-0.168×10^6	-0.362×10^6	-0.403×10^6
c	-1.193×10^4	-0.687×10^4	0.045×10^4
d	3.1	3.6	3.9
	B3LYP/G3large		
a	1.488	2.122	3.734
b	-0.145×10^6	-0.258×10^6	-0.318×10^6
c	-1.174×10^4	-0.674×10^4	-0.025×10^4
d	2.9	3.3	3.6

this constraint are given in Table 3. Comparison to Table 2 shows that the *a*, *b*, and *c* parameters become somewhat different under this constraint, changing in some instances by more than a factor of 2. This explains the failure of an attempt to find universal values of all four parameters for all of the QM methods considered (data not shown), which led to very large MUEs at each cavity size.

The results reported in the tables correspond to optimizing the FESR parameters over the class of all solutes taken together. At the expense of losing generality, even better results could be obtained if the parameters were to be optimized separately for each of the subclasses of neutrals, cations, and anions. For example, with the B3LYP/6-31+G* method taken with the ρ_0 = 0.001 au cavity size, such separate optimizations lead to MUE of 0.85, 1.61, and 2.70 kcal/mol for neutrals, cations, and anions, respectively. This corresponds to improvements of about 0.1−0.2 kcal/mol for these subclasses over the results in Table 1, where the parameters take on more compromise values from optimization over all classes.

To get an idea of the extent to which the parameters and results might depend on the database, the 372 solutes were divided into two groups by alternately assigning each solute in the given list to either a training set or a test set. B3LYP/6-31+G* calculations with the ρ_0 = 0.001 au cavity size together with optimization of parameters only over solutes in the training set led to a change in the MUE of the test set by only 0.02 kcal/mol from that of training over the entire database. The fitting parameters *a*, *b*, and *c* changed by less than 10%, and *d* changed not at all. A second analogous test made by interchanging the roles of the training and test sets produced essentially the same results. This exercise demonstrates the stability and robustness of the parametrization and further indicates that the model should have a useful predictive power for solvation energies of molecules not contained in the data set.

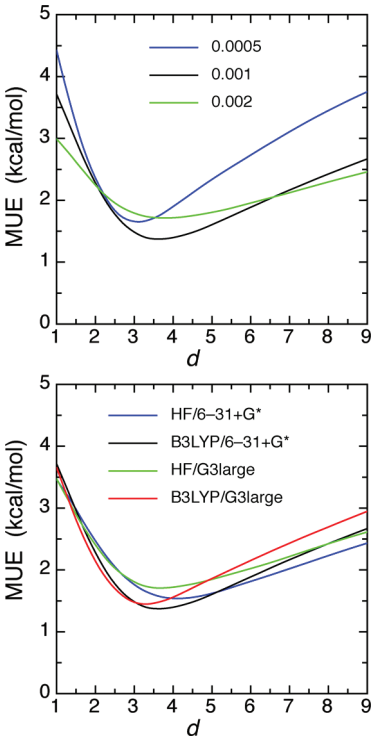

Figure 3. Mean unsigned error as a function of exponent *d*. The top panel shows results for the B3LYP/6-31+G* method with several cavity sizes. The bottom panel shows results for the ρ_0 = 0.001 au cavity size with several QM methods.

Dependence on Cavity Size. Most of the trends with cavity size have already been discussed above in connection with other related matters. It remains here to explicitly note that the best SS(V)PE+FESR MUE results are obtained in all cases examined with the contour ρ_0 = 0.001 au, which is the most commonly used value and has also proved to be near optimal in previous studies of dielectric effects in neutrals.[60,61] In contrast, with SS(V)PE alone, the lowest MUE for neutrals occurs with ρ= 0.0005 au, for cations with ρ = 0.001 au, and for anions with ρ = 0.002 au.

Among the various classes of solutes, cations are the most sensitive to the cavity size. The change in SS(V)PE+FESR MUE with each step in cavity size is about 1 kcal/mol for cations, whereas this change is about 0.3 kcal/mol for neutrals and about 0.2 kcal/mol for anions.

Behavior of Selected Solutes. Here, we describe some particularly notable behaviors of certain individual solutes, mainly to indicate areas where additional improvements can be sought in future work. To keep the discussion manageable, we discuss only the case of B3LYP/6-31+G* calculations with the ρ_0 = 0.001 au contour and optimization of FESR parameters over all classes of solutes. Other computational protocols show similar behavior.

The worst numerical result among the neutrals is for N,N′-dimethylpiperazine, where the experimental hydration energy is −7.58 kcal/mol while the SS(V)PE+FESR result is −2.17 kcal/mol, corresponding to an error of 5.41 kcal/mol. The FESR result in this case is actually slightly worse than the SS(V)PE result, which is −2.86 kcal/mol, corresponding to an error of 4.72 kcal/mol for this compound.

Table 3. Optimum Parameter Values Obtained from Several QM Methods and Cavity Sizes, under the Constraint of All QM Methods Sharing a Common Universal Value of *d* at Each Cavity Size

parameter	ρ_0 = 0.0005 au	ρ_0 = 0.001 au	ρ_0 = 0.002 au
d	3.1	3.6	3.8
	HF/6-31+G*		
a	2.283	3.221	5.308
b	-0.162×10^6	-0.343×10^6	-0.293×10^6
c	-1.101×10^4	-0.756×10^4	0.055×10^4
	B3LYP/6-31+G*		
a	1.570	2.345	4.158
b	-0.259×10^6	-0.585×10^6	-0.523×10^6
c	-1.543×10^4	-1.107×10^4	0.027×10^4
	HF/G3large		
a	1.157	2.094	4.284
b	-0.168×10^6	-0.362×10^6	-0.311×10^6
c	-1.193×10^4	-0.687×10^4	0.041×10^4
	B3LYP/G3large		
a	1.119	1.788	3.623
b	-0.260×10^6	-0.598×10^6	-0.544×10^6
c	-1.794×10^4	-1.171×10^4	-0.036×10^4

Being quite nonpolar, the 41 neutral unsubstituted hydrocarbons in the database generally have very small experimental hydration energies, in some cases even being positive, indicating hydrophobicity. By itself, SS(V)PE necessarily produces negative hydration energies, and it therefore gives a poor description of these solutes. The electric fields on the cavity surface of such solutes are generally quite small, consistent with essentially no hydrogen bonding to water, so the field-dependent terms in the FESR model are also ineffective in describing these solutes. In particular, the field-dependent terms contribute only about −0.01 kcal/mol for alkanes, about −0.07 kcal/mol for alkenes, about −0.25 kcal/mol for alkynes, and about −0.02 kcal/mol for aromatic hydrocarbons. The *a* term in the FESR model then becomes the most significant contribution, and its positive value is mainly responsible for the improvement in description of the unsubstituted hydrocarbons. The neutral unsubstituted hydrocarbons are all clustered in the top right corners of the panels in Figure 2, where it is seen that the FESR correction does not noticeably change the pattern for these solutes, except for providing a constant horizontal shift from the *a* term. Similar conclusions regarding an analogous positive constant term were drawn in a previous study of alkane hydration energies.[81]

For neutral halocarbons, the field dependent terms provide free energy contributions ranging from −0.03 to −0.37 kcal/mol, such that the total FESR correction is 1.97 to 2.31 kcal/mol. The large range in experimental free energies for this class of compounds (from −2.73 kcal/mol for bromotoluene up to 4.28 kcal/mol for octafluoropropane) is already mostly accounted for in the SS(V)PE results, while the constant shift from the FESR *a* term helps to reduce the MUE for this subclass from 2.38 to 0.68 kcal/mol.

Among cations, the errors in hydration energies of protonated 4-nitroaniline, 3-aminoaniline, ammonia, and dimethyl sulfide are increased by more than 1 kcal/mol upon FESR correction, and the errors are found to be quite sensitive to computation level and cavity setting. All other cation results are improved by the FESR correction.

All anions improve their hydration energies upon applying the FESR correction. The largest anion error occurs for F^-, not surprisingly in light of its extremely high concentration of negative charge into a very small volume, where the error is reduced from the huge −21.3 kcal/mol at the SS(V)PE level to the still large 13.4 kcal/mol after FESR correction. However, in the case of OH^-, which is the other very small anion in the database, the similarly huge SS(V)PE error of −26.8 kcal/mol is quite effectively reduced by the FESR correction to only 3.7 kcal/mol.

CONCLUDING REMARKS

The SS(V)PE+FESR model to calculate hydration free energy was introduced in a previous short communication[31] that demonstrated its potential utility and efficiency. The present contribution studies the performance of this model under a variety of computational conditions and also gives additional information on its behavior in various respects.

While all of the QM methods considered give good performance in this application, there are some differences worthy of note. Inclusion of electron correlation in the solute QM method, as studied by comparison of B3LYP to HF calculations, leads overall to better results. However, extension of the basis set in the solute QM method, as studied by comparison of the G3large to the modest 6-31+G* basis set, generally leads to slightly poorer results. The cavity obtained from the ρ_0 = 0.001 au contour gives significantly better results than with either larger or smaller cavities from 0.0005 or 0.002 au contours, respectively.

Overall, the best SS(V)PE+FESR results in this work are obtained with the B3LYP/6-31+G* QM method and the 0.001 au cavity contour. This gives a MUE for neutrals of 0.9 kcal/mol, for cations of 1.8 kcal/mol, for anions of 2.9 kcal/mol, and, combining the latter two, for all ions of 2.4 kcal/mol. This performance is comparable to the best analogous literature results using the SMVLE method[21] that was trained over almost the same database as used here and that used the SVPE dielectric continuum model, which should give essentially the same results as SS(V)PE for the solutes concerned. After optimization of a large number of parameters, including 20 atom-specific surface tension coefficients and 12 generic parameters involving the solute electric field on the cavity surface, the SMVLE model achieved a MUE of 0.5 kcal/mol for the neutrals and 3.1 kcal/mol for the ions. Compared to this, the best SS(V)PE+FESR results are not quite as good for neutrals and are better for ions, despite the much smaller number of adjustable parameters. Our view is that minimizing the number of empirical parameters in a model increases the probability that the parameters will have physical significance that can be used to guide development of more fundamentally based theories.

The FESR correction utilizes a functional form where hydration free energies have a nonlinear dependence on the minimum and maximum values of the solute normal electric field on the cavity surface. It is argued that these field-dependent terms mainly describe specific hydrogen bonding interactions between solute and solvent. The nonlinear parameter *d* of the model is always found to be in the range of 2.9−4.1 with all of the combinations of computational protocols considered, its more exact value depending to some extent on QM method and cavity size. In fact, nearly universal values of *d* exist for each cavity size independent of the QM method used. Furthermore, negligible

improvement in the MUE is achieved by allowing cations and anions to have different values of *d*. These observations suggest some real underlying physical significance of the nonlinear *d* exponent of the field extrema for describing hydrogen bonding energies.

The SS(V)PE+FESR model also invokes a constant term *a* that significantly improves the hydration energies of nonpolar solutes and, by extension, of other solutes as well. This parameter likely describes in some rough overall average sense the short-range contributions from cavitation, exchange repulsion, and dispersion. While such interactions are actually size-dependent, no significant effect was found on including additional terms proportional to solute surface area or volume. This may be due to an insufficient range of sizes in the database used to train the parameters.

Further research will focus on more specifically modeling short-range contributions from cavitation, exchange repulsion, and dispersion, which can be expected to particularly improve the treatment of neutral solutes. Improvement in the treatment of ions should also follow, but that may be difficult to discern in practice because the best SS(V)PE+FESR results for ions are already in better agreement with experimental results than the estimated experimental error.

It is concluded that the FESR correction is quite effective in removing a large part of the error in hydration energies from the SS(V)PE model. There is every reason to believe that an analogous FESR correction would perform similarly well as an adjunct to other dielectric continuum models in common use. With the parameters reported in this work, the SS(V)PE+FESR model can be used to predict hydration energies of solutes not in the database used for its training. Additionally, it should also serve as a valuable building block for further development of models to bring results from implicit solvation calculations into even closer agreement with experimental results.

■ ASSOCIATED CONTENT

Supporting Information. Tables of compounds that failed the surface integration for at least one QM protocol (S1); compounds that required a nondefault surface integration origin (S2); and listings for each solute under each QM protocol of the electric field extrema, hydration free energies, and related errors (S3, S4, S5, S6). This material is available free of charge via the Internet at http://pubs.acs.org.proxy.library.nd.edu.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: chipman.1@nd.edu.

■ ACKNOWLEDGMENT

The generosity of the Minnesota group in making available their extensive Solvation Database is gratefully acknowledged. This material is based upon work supported by the Department of Energy under Award Number DE-SC0002216. This is Contribution No. NDRL-4897 from the Notre Dame Radiation Laboratory.

■ REFERENCES

(1) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
(2) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
(3) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.
(4) Miertuš, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
(5) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(6) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.
(7) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775.
(8) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.
(9) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074.
(10) Amovilli, C.; Barone, V.; Cammi, R.; Cancès, E.; Cossi, M.; Mennucci, B.; Pomelli, C.; Tomasi, J. *Adv. Quant. Chem.* **1999**, *32*, 227.
(11) Pliego, J. R.; Riveros, J. M. *J. Phys. Chem. A* **2001**, *105*, 7241.
(12) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.
(13) Piquemal, J. P.; Marquez, A.; Parisel, O.; Giessner-Prettre, C. *J. Comput. Chem.* **2005**, *26*, 1052.
(14) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 16066.
(15) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. *J. Chem. Theory Comput.* **2006**, *2*, 128.
(16) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2011.
(17) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760.
(18) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J. *Acc. Chem. Res.* **2009**, *42*, 489.
(19) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2009**, *42*, 493.
(20) Zuo, C.-S.; Wiest, O.; Wu, Y.-D. *J. Phys. Chem. A* **2009**, *113*, 12028.
(21) Liu, J.; Kelly, C. P.; Goren, A. C.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G.; Zhan, C.-G. *J. Chem. Theory Comput.* **2010**, *6*, 1109.
(22) Umeyama, H.; Morokuma, K. *J. Am. Chem. Soc.* **1977**, *99*, 1316.
(23) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066.
(24) Stevens, W. J.; Fink, W. H. *Chem. Phys. Lett.* **1987**, *139*, 15.
(25) Glendening, E. D.; Streitweiser, A. *J. Chem. Phys.* **1994**, *100*, 2900.
(26) Chen, W.; Gordon, M. S. *J. Phys. Chem.* **1996**, *100*, 14316.
(27) Isaacs, E. D.; Shukla, A.; Platzman, P. M.; Hamann, D. R.; Barbiellini, B.; Tulk, C. A. *Phys. Rev. Lett.* **1999**, *82*, 600.
(28) Ghanty, T. K.; Staroverov, V. N.; Koren, P. R.; Davidson, E. R. *J. Am. Chem. Soc.* **2000**, *122*, 1210.
(29) Barbiellini, B.; Shukla, A. *Phys. Rev. B* **2002**, *66*, 235101.
(30) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. *Chem.—Eur. J.* **2009**, *15*, 851.
(31) Pomogaeva, A.; Thompson, D. W.; Chipman, D. M. *Chem. Phys. Lett.* **2011**, *511*, 161.
(32) Sadlej, J.; Buch, V.; Kazimirski, J. K.; Buck, U. *J. Phys. Chem. A* **1999**, *103*, 4933.
(33) Chipman, D. M. *J. Chem. Phys.* **2003**, *118*, 9937.
(34) Chipman, D. M.; Chen, F. W. *J. Chem. Phys.* **2006**, *124*, 144507.
(35) 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 2009; University of Minnesota: Minneapolis, MN, 2009.
(36) Ben-Naim, A. *J. Phys. Chem.* **1978**, *82*, 792.
(37) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. *J. Phys. Chem. A* **1998**, *102*, 7787.
(38) Ho, J.; Klamt, A.; Coote, M. L. *J. Phys. Chem. A* **2010**, *114*, 13442.
(39) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(40) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(41) We use the B3LYP version implementing the VWN5 electron gas correlation formula.
(42) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.
(43) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
(44) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. *J. Comput. Chem.* **1983**, *4*, 294.

(45) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.
(46) Luque, F. J.; Orozco, M.; Bhadane, P. K.; Gadre, S. R. *J. Chem. Phys.* **1994**, *100*, 6718.
(47) Cabral do Couto, P. C.; Chipman, D. M. *J. Phys. Chem. A* **2010**, *114*, 12788.
(48) Luque, F. J.; Gadre, S. R.; Bhadane, P. K.; Orozco, M. *Chem. Phys. Lett.* **1995**, *232*, 509.
(49) Chipman, D. M. *J. Chem. Phys.* **1999**, *110*, 8012.
(50) Chipman, D. M. *J. Chem. Phys.* **2000**, *112*, 5558.
(51) Chipman, D. M. *J. Chem. Phys.* **1997**, *106*, 10194.
(52) Zhan, C.-G.; Bentley, J.; Chipman, D. M. *J. Chem. Phys.* **1998**, *108*, 177.
(53) Chipman, D. M. *J. Chem. Phys.* **2006**, *124*, 224111.
(54) Vilkas, M. J.; Zhan, C.-G. *J. Chem. Phys.* **2008**, *129*, 194109.
(55) Amovilli, C.; Filippi, C.; Floris, F. M. *J. Chem. Phys.* **2008**, *129*, 244106.
(56) Cancès, E.; Mennucci, B. *J. Chem. Phys.* **2001**, *114*, 4744.
(57) Chipman, D. M. *Theor. Chem. Acc.* **2002**, *107*, 80.
(58) Mennucci, B.; Cammi, R.; Tomasi, J. *J. Chem. Phys.* **1998**, *109*, 2798.
(59) Cossi, M.; Barone, V. *J. Phys. Chem. A* **2000**, *104*, 10614.
(60) Zhan, C.-G.; Chipman, D. M. *J. Chem. Phys.* **1998**, *109*, 10543.
(61) Zhan, C.-G.; Chipman, D. M. *J. Chem. Phys.* **1999**, *110*, 1611.
(62) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616.
(63) Floris, F. M.; Tomasi, J.; Pascual-Ahuir, J. L. *J. Comput. Chem.* **1991**, *12*, 784.
(64) Rösch, N.; Zerner, M. C. *J. Phys. Chem.* **1994**, *98*, 5817.
(65) Olivares del Valle, F. J.; Aguilar, M. A. *THEOCHEM* **1993**, *99*, 25.
(66) Amovilli, C. *Chem. Phys. Lett.* **1994**, *229*, 244.
(67) Amovilli, C.; Mennucci, B. *J. Phys. Chem. B* **1997**, *101*, 1051.
(68) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
(69) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.
(70) Smith, R.; Tanford, C. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 289.
(71) Chothia, C. *Nature* **1974**, *248*, 338.
(72) Reynolds, J. A.; Gilbert, D. B.; Tanford, C. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 2925.
(73) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717.
(74) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. *Science* **1991**, *252*, 106.
(75) Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. *J. Am. Chem. Soc.* **1999**, *121*, 9243.
(76) Huang, D. M.; Chandler, D. *J. Phys. Chem. B* **2002**, *106*, 2047.
(77) Chandler, D. *Nature* **2005**, *437*, 640.
(78) Floris, F. M.; Selmi, M.; Tani, A.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 6353.
(79) Amovilli, C.; Floris, F. M. *Phys. Chem. Chem. Phys.* **2003**, *5*, 363.
(80) Ashbaugh, H. S.; Pratt, L. R. *Rev. Mod. Phys.* **2006**, *78*, 159.
(81) Carlson, H. A.; Jorgensen, W. L. *J. Phys. Chem.* **1995**, *99*, 10667.