

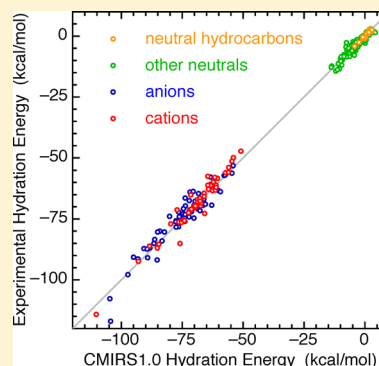
Hydration Energy from a Composite Method for Implicit Representation of Solvent

Anna Pomogaeva[†] and Daniel M. Chipman*

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

S Supporting Information

ABSTRACT: The CMIRS1.0 (composite method for implicit representation of solvent, Version 1.0) model is introduced for efficacious and inexpensive computation of hydration free energies. The method collects together several disparate models designed to describe short-range dispersion, exchange, and hydrogen bonding interactions as well as long-range electrostatic interactions. All the interactions are formulated as functionals of the solute charge density. The model uses only six adjustable parameters to determine the various short-range terms. In conjunction with an isodensity criterion that uses one parameter to determine the solute cavity size and shape, the model is tested on a large database of neutral and ionic solutes in water. The mean unsigned error compared to experiment is found to be as low as 0.8 kcal/mol for neutral solutes and 2.4 kcal/mol for ionic solutes, which is comparable to or better than other analogous approaches in the literature that invoke many more fitting parameters.



INTRODUCTION

Being a universal medium for accommodation of a wide range of chemical and biological species and transformations, water is the most important solvent. The change in energy when a molecule is transferred from the gas phase into an aqueous environment is a particularly important quantity that modulates thermodynamic stability, pK_a , redox potential, and various other important properties. Computational difficulties concerning this issue are well-known. Bulk solvent acts as a dielectric medium; thus it is essential to treat the long-range electrostatic interactions between solute and water. Various short-range interactions are also significant. For example, a solute will engage in dispersion attraction and exchange repulsion interactions with nearby water molecules. The possibility of forming solute–solvent hydrogen bonds is also relevant, especially for ionic solutes. Computation of such interactions between a solute and a sufficient number of explicitly treated water molecules is expensive, which has led to considerable activity in development of efficient implicit models of solvation. Several excellent reviews are available on various methods developed for this purpose, discussing their relative merits, progress, and achievements.^{1–5}

For polar solutes in polar solvents the solvation free energy is generally dominated by the long-range electrostatic interactions with bulk solvent arising from solvent polarization by the solute, and back-polarization of the solute by the polarized solvent can also be significant. Implicit solvation models for the long-range electrostatic interactions are usually evaluated with dielectric continuum models based on solving the Poisson equation^{6,7} for the electrostatic potential produced by the solute charge density at the surface of a cavity nominally containing the solute, or alternatively with an approximate analytic solution known as the generalized Born equation.^{8,9}

Short-range interactions are generally nonnegligible and may be dominant if the solvent and/or the solute is nonpolar. Many workers evaluate the entire solvation energy with dielectric continuum models by adjusting parameters, such as the atomic radii, to produce the best agreement with experimental data sets, thereby indirectly incorporating the energetic effects of short-range interactions in some average manner. This strategy is most successful with a collection of similar solutes and is less so with solutes having a wide range of electronic structures.

In other implicit solvation approaches that have the potential for more general applicability, additional models of the short-range interactions are used in conjunction with the dielectric continuum models. These additional models usually focus on dispersion attraction and exchange repulsion, which are believed to be separately large in magnitude but largely canceling to produce a smaller net result. The simplest and most common of such models attempts to treat only the combined effect of dispersion and exchange by asserting a proportionality to the solvent accessible portions of the surface areas of those solute atoms that are partially in contact with solvent.^{2,8} Several previous approaches have developed more sophisticated separate models of dispersion and exchange. In our previous work¹⁰ we provided a discussion of dispersion models based on a reaction field approach,^{11–16} on configuration interaction calculations,^{17–20} or on use of empirical atom-centered potentials,^{21–23} and of exchange models based on use of empirical atom-centered potentials²² or on asserting a proportionality to the amount of solute charge density penetrating outside the cavity nominally containing it.¹⁹ To this list should be added some more recent and promising

Received: October 14, 2013

Published: November 14, 2013



developments. One development treats dispersion in terms of the spherically averaged static polarizability of the solute and the refractive index of the solvent,²⁴ whereas another development treats dispersion in terms of the frequency-dependent multipolar polarizabilities of the solute and the dielectric function of water at imaginary frequencies.^{25,26} Our own recent work¹⁰ has introduced other new models of dispersion and exchange described in more detail below that proved promising in test calculations of neutral solutes in the nonpolar solvents cyclohexane and benzene.

In water solvent hydrogen bonding interactions are also often important. In principle, sufficiently accurate treatment of electrostatics, dispersion, exchange, and possibly also charge transfer should automatically describe hydrogen bonding interactions. However, hydrogen bonding is a very specific and local effect that will probably elude description by the kinds of generic implicit solvation models described above, so it is meet to also include an additional model for solute–solvent hydrogen bonding. The COSMO-RS method²⁷ treats hydrogen bonding by a novel statistical mechanical analysis using the apparent surface charge density produced by the COSMO dielectric continuum method. We have previously^{28,29} introduced a method called the FESR (field-extremum short-range) model that describes hydrogen bonding in terms of nonlinear dependencies on the minima and maxima of the electric field produced by the solute at the cavity surface. The method also invoked a constant term to describe the contributions from dispersion and exchange in some overall average manner. Test calculations on neutral and ionic solutes in water showed that inclusion of the FESR model gave significant improvement over the dielectric continuum treatment alone, particularly for ions.

In this work we combine our previous developments for short-range interactions^{10,28,29} into the DEFESR (dispersion, exchange, and field-extremum short-range) model. This is to be used in conjunction with the SS(V)PE (surface and simulation of volume polarization for electrostatics)^{30–33} dielectric continuum model. For conciseness, the SS(V)PE and current DEFESR models taken together are dubbed as the CMIRS1.0 (composite method for implicit representation of solvent, Version 1.0) model. The version number is given in anticipation of future improvements, particularly to various aspects of the DEFESR contribution. A more detailed description of the various components included in CMIRS1.0 are provided in the next section. The following section provides details on the computational methods used to illustrate the accuracy of CMIRS1.0 for water solvent. The next section describes the optimization of the small number of parameters involved in the model and the following section discusses numerical results for a collection of neutral and ionic solutes in water. A brief conclusion section ends the paper.

■ CMIRS1.0 MODEL

In the CMIRS1.0 model the experimental free energy of solvation is approximated by

$$\Delta G_{\text{CMIRS1.0}} = \Delta G_{\text{SS(V)PE}} + \Delta G_{\text{DEFESR}} \quad (1)$$

where $\Delta G_{\text{SS(V)PE}}$ describes the long-range electrostatic interaction with bulk solvent and ΔG_{DEFESR} collects the various short-range interactions. The SS(V)PE dielectric continuum model is an approximation to the SVPE (surface and volume polarization for electrostatics) model^{30,34–37} that provides an exact treatment of the electrostatic effect arising from the solute charge density that penetrates into the volume outside the

cavity nominally containing the solute. If implemented with the same cavity, the SS(V)PE method becomes equivalent^{33,38} to the modified version of the IEF-PCM method that is often currently used.^{39,40} The DEFESR contribution is given as

$$\Delta G_{\text{DEFESR}} = \Delta G_{\text{Disp}} + \Delta G_{\text{Exch}} + \Delta G_{\text{FE}} \quad (2)$$

which terms are individually briefly described in the following subsections.

All terms in CMIRS1.0 are formulated as functionals of the solute charge density, which is determined by quantum mechanical (QM) calculation of the solute wave function. The reaction potential generated by the SS(V)PE contribution is allowed to back-polarize the solute, and the solvent and solute polarizations are iterated until mutual self-consistency is reached. In the present 1.0 version of the model the DEFESR contributions are evaluated only once from the final solute charge density. One avenue for future improvements to the CMIRS1.0 model will allow the DEFESR contributions to also polarize the solute.

Dispersion Interactions. Our dispersion model is adapted from the VV09 nonlocal density functional originally suggested by Vydrov and Van Voorhis⁴¹ to account for dispersion interactions between atoms and molecules in the gas phase. The adaptation to an implicit solvation model involves identifying one of the interacting partners with the solute whose computed electronic charge density is denoted $\rho(\mathbf{r})$ and the other partner with the structureless solvent whose average electronic charge density $\bar{\rho}_{\text{solvent}}$ is taken from experiment. The working formula in atomic units (where \hbar , e , m_e , and $1/4\pi\epsilon_0$ each have a value of unity) is

$$\Delta G_{\text{Disp}} = A \int_{\text{solvent}} d^3\mathbf{r} \times \frac{\rho(\mathbf{r})}{\sqrt{\rho(\mathbf{r}) + \frac{3C}{4\pi} \left| \frac{\nabla \rho(\mathbf{r})}{\rho(\mathbf{r})} \right|^4} \left[\sqrt{\rho(\mathbf{r}) + \frac{3C}{4\pi} \left| \frac{\nabla \rho(\mathbf{r})}{\rho(\mathbf{r})} \right|^4} + \sqrt{\bar{\rho}_{\text{solvent}}} \right]} I(\mathbf{r}; \delta)$$

where

$$I(\mathbf{r}; \delta) = \int_{\text{solvent}} d^3\mathbf{r}' \frac{1}{|\mathbf{r} - \mathbf{r}'|^6 + \delta^6}$$

Here C is a constant taken over unchanged from the value 0.0089 that was optimized to fit a test set of gas-phase intermolecular interaction energies.⁴² The linear proportionality constant A as well as the dispersion damping parameter δ introduced to prevent dispersion catastrophe at very short distances are parameters to be optimized. A linear factor of $(\bar{\rho}_{\text{solvent}})^{1/2}$ that would otherwise appear is absorbed into the A parameter for convenience.

The implementation to evaluate the necessary integrals for ΔG_{Disp} is the same as that described previously.¹⁰ It is the most demanding part of the DEFESR calculation, approximately doubling the total time compared to SS(V)PE alone for large solutes and costing relatively even more with small solutes.

We also consider an alternative formula based on setting $\nabla \rho(\mathbf{r})$ to zero, which can be viewed as arising from a simpler nonlocal density functional developed by Andersson et al.⁴³

Exchange Interactions. Our exchange model is adapted from an approach that represents the asymptotic exchange energy between two gas-phase one-electron atoms as a surface integral over the flux of exchanging electrons.^{44–46} The working formula is

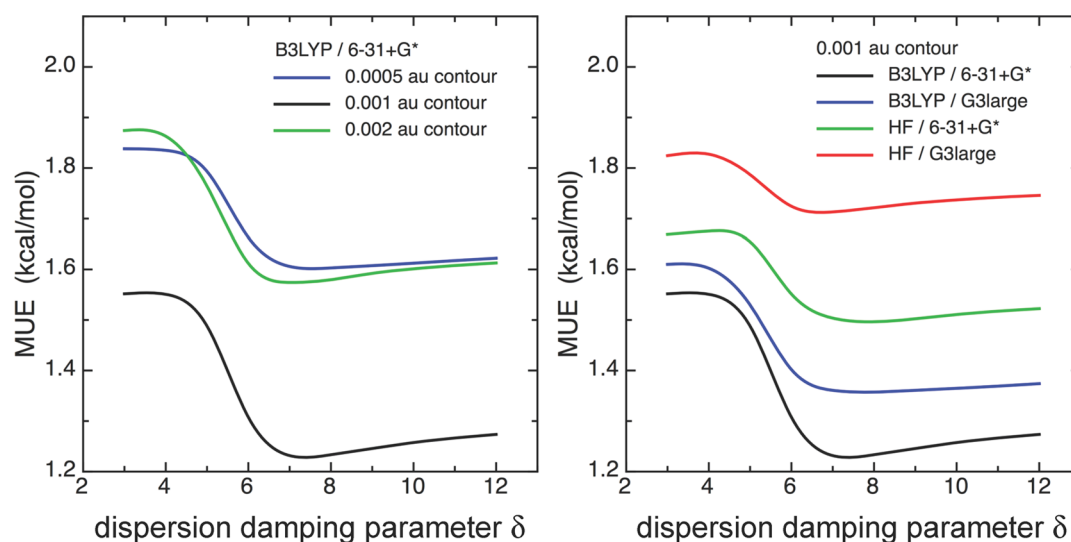


Figure 1. Mean unsigned errors over all solutes as a function of the dispersion damping factor δ .

$$\Delta G_{\text{Exch}} = B \int_{\text{solvent}} d^3\mathbf{r} |\nabla \rho(\mathbf{r})|$$

where the linear proportionality constant B is a parameter to be optimized. A linear factor of $\bar{\rho}_{\text{solvent}}$ that would otherwise appear is absorbed into the B parameter for convenience. The implementation to evaluate the necessary integrals for ΔG_{Exch} is the same as that described previously.¹⁰ We note that unconstrained QM calculation of the solute charge density will necessarily produce a tail that penetrates into the solvent region, thereby allowing ΔG_{Exch} to be nonzero.

We also consider an alternative formula based on replacing $|\nabla \rho(\mathbf{r})|$ with $\rho(\mathbf{r})$, as has been used previously.¹⁹

Hydrogen Bonding Interactions. Motivated by earlier studies indicating that the strength of a hydrogen bond is related to the strength of the electric field in its vicinity^{47–49} our previous FESR development^{28,29} quantified the relation. Dropping the additive constant invoked there, the current working formula for the Field-Extremum contribution describing hydrogen bonding is

$$\Delta G_{\text{FE}} = C |F_{\text{min}}|^{\gamma} + D F_{\text{max}}^{\gamma}$$

which depends on the minimum F_{min} and maximum F_{max} values of the outgoing normal electric field produced by the solute on the cavity surface. If the solute has no negative values of F_{min} anywhere on the cavity surface, then its F_{min} is assigned the value zero. Similarly, if a solute has no positive values of F_{max} , then its F_{max} is assigned the value zero. The linear proportionality constants C and D as well as the exponent γ are parameters to be optimized. Our previous study with the SS(V)PE+FESR model^{28,29} found no significant improvement in allowing the two terms to have different exponents γ , suggesting some underlying physical significance for the common value. We therefore assume equal exponents in the two terms here as well.

■ COMPUTATIONAL DETAILS

In the present research the experimental hydration free energies used for the parametrization procedure and for evaluation of the accuracy were taken from the Minnesota Solvation Database—Version 2009⁵⁰ that has been collected and kindly provided as an open source by the Minnesota group. The

standard state assumed there and here corresponds to the Ben-Naim convention⁵¹ of identical concentrations in gas and in solution. Ionic hydration free energies are based on the Tissandier et al.⁵² value of -265.9 kcal/mol for the proton hydration free energy. Uncertainties in the data are stated⁵⁰ to be about 0.2 kcal/mol for neutrals and about 3 kcal/mol for ions. The database also includes optimized gas-phase geometries for all solutes. We adopted these geometries in our work as unchanged upon solvation. The solutes are generally all fairly rigid molecules, except for free methyl rotors, so it is believed that changes in geometry and zero-point vibration energy due to solvation will not significantly affect the results.

To investigate the effects of electron correlation in the solute, the QM calculations of solute charge density were carried out both with uncorrelated Hartree–Fock (HF) ab initio methods and with correlated B3LYP^{53,54} density functional methods. The latter utilized the version implementing the VWN5 electron gas correlation formula.

The influence of the basis set was considered with the modest-sized 6-31+G*^{55–57} basis and with the rather larger G3large⁵⁸ basis set. The former is of double- ζ quality in the valence space and has polarization as well as diffuse functions on the heavy atoms. The latter is of triple- ζ quality in the valence space and has multiple polarization as well as diffuse functions on all atoms.

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 that nominally encapsulates the solute molecule and 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 so 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 $\rho_0 = 0.0005$, 0.001, and 0.002 au which were found to be reasonable in previous studies.^{59,60}

Unless otherwise specified, the DEFESR linear parameters A , B , C , and D and nonlinear parameters δ and γ were adjusted to minimize the least-squares deviation from experiment, giving equal weight to all solutes in the chosen training set.

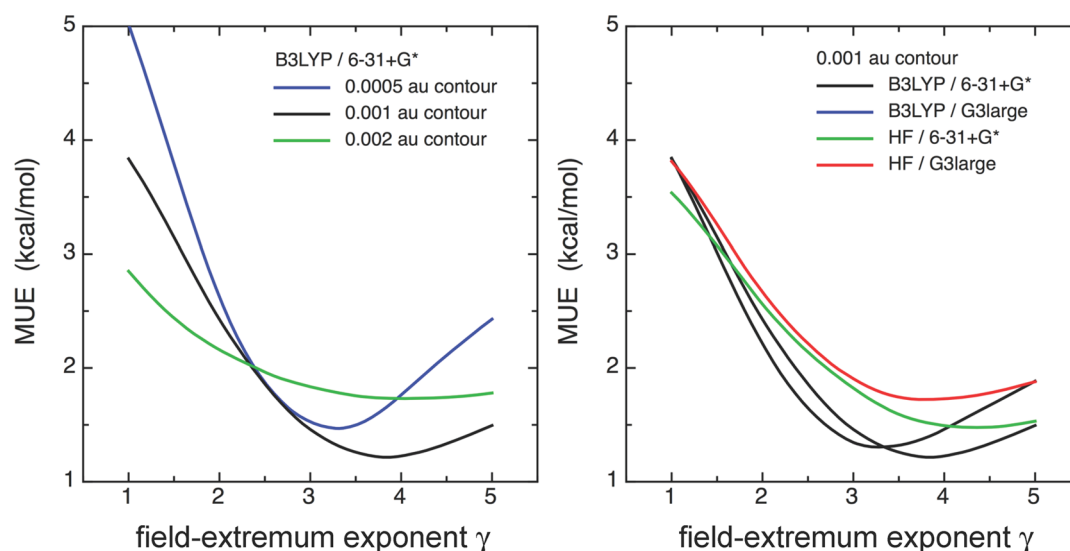


Figure 2. Mean unsigned errors over all solutes as a function of the field-extremum exponent γ .

Table 1. Optimum Values of Linear Parameters,^a All in au

QM method	ρ_0^b	N^c	A^d	B^e	C^f	D^g
B3LYP/6-31+G*	0.0005	376	−0.009663	0.063101	−1840.0	−70.873
	0.001	376	−0.010825	0.045576	−944.4	−17.817
	0.002	359	−0.011463	0.031049	−482.3	0.389
B3LYP/G3large	0.0005	376	−0.009069	0.057990	−1873.9	−86.321
	0.001	374	−0.010186	0.042081	−980.4	−19.923
	0.002	357	−0.010765	0.028813	−506.0	−0.446
HF/6-31+G*	0.0005	374	−0.007124	0.052642	−1069.8	−47.719
	0.001	372	−0.008314	0.039413	−544.9	−11.378
	0.002	357	−0.009096	0.027865	−269.8	1.025
HF/G3large	0.0005	375	−0.007746	0.054009	−1186.1	−57.669
	0.001	371	−0.006589	0.032809	−620.5	−12.664
	0.002	357	−0.009262	0.027527	−303.0	0.330

^aThe nonlinear parameters are fixed at the values $\delta = 7$ au and $\gamma = 3.6$. ^bCavity size isodensity contour in au. ^cNumber of solutes used for optimization. ^dDispersion proportionality parameter in au. ^eExchange proportionality parameter in au. ^{f $|F_{\min}|^6$ proportionality parameter in au. ^g F_{\max}^{ν} proportionality parameter in au.}

Experimental properties for bulk water were taken as 78.36 for the dielectric constant and 0.05 au for $\bar{\rho}_{\text{solvent}}$.

Our implementation of the SS(V)PE contribution utilizes a single-center method to carry out integrations over the solute cavity surface, usually with 1202 Lebedev quadrature points, or more in cases where diagnostics indicated that more were required. In a few cases this procedure failed because of severely nonspherical cavity shapes, most often only at the smallest cavity size corresponding to the 0.002 au contour value. Thus, of the 274 neutral solutes in the database, 10 failed in all the calculations and 18 more failed for at least one QM method at one cavity contour value. Of the 52 cations in the database 7 failed for at least one QM method at one cavity contour value. All 60 anions in the database were treated successfully. Thus, the parametrization and evaluation was based, depending on the QM method and the cavity contour value, on 249–264 neutrals, 48–52 cations, 60 anions, and 357–376 total solutes. It is believed that the resulting slight differences in sizes of the data sets should not significantly bias any of the computational methods from being fairly compared to the others.

PARAMETER OPTIMIZATION

Figure 1 presents results for the mean unsigned error (MUE) of the calculated hydration energies of all solutes as a function of the dispersion damping factor δ . In each case the linear parameters A , B , C , and D were reoptimized at each value of δ , while the nonlinear parameter γ was held fixed at the value 3.6. The left panel gives results for different cavity sizes with the B3LYP/6-31+G* method, where the optimum value of δ varies from 7.1 to 7.5 au. The right panel gives results for different QM methods with the cavity size contour 0.001 au, where the optimum value of δ varies from 6.7 to 7.9 au. Our previous work¹⁰ where this parameter was optimized for a collection of neutral solutes in cyclohexane and benzene solvents found optimum values in the range 6–8 au, depending on the solvent and the cavity size. It therefore appears that the value 7 au (i.e., about 3.7 Å) is nearly optimal for various cavity sizes, QM methods, and solvents, and that value will be adopted for δ throughout the remainder of this paper. It is interesting that similar optimum values have also been found in a study of a large collection of gas-phase noncovalent intermolecular energies evaluated with HF and various density functionals.⁶²

Figure 2 presents results for the MUE of the calculated hydration energies as a function of the field-extremum

Table 2. Mean Unsigned Errors Compared to Experiment, in kcal/mol

QM method	ρ_0^a	hydrocarbons	all neutrals	cations	anions	all ions	all solutes
B3LYP/6-31+G*	0.0005	0.94	0.99	3.12	3.01	3.06	1.61
	0.001	0.36	0.75	1.94	2.73	2.36	1.23
	0.002	1.11	0.95	3.10	2.87	2.97	1.57
B3LYP/G3large	0.0005	1.43	1.22	3.31	3.10	3.20	1.81
	0.001	0.59	0.89	1.98	2.89	2.47	1.36
	0.002	0.76	0.88	2.83	2.98	2.91	1.49
HF/6-31+G*	0.0005	0.54	1.09	3.28	2.87	3.06	1.68
	0.001	0.66	1.10	2.32	2.57	2.45	1.50
	0.002	1.77	1.51	3.39	2.60	2.95	1.95
HF/G3large	0.0005	1.03	1.23	3.45	2.88	3.14	1.80
	0.001	0.64	1.37	2.40	2.62	2.52	1.71
	0.002	1.43	1.39	3.22	2.58	2.86	1.84

^aCavity size isodensity contour in au.

exponent γ . In each case the linear parameters A , B , C , and D were reoptimized at each value of γ , while the nonlinear parameter δ was held fixed at the value 7 au. The left panel gives results for different cavity sizes with the B3LYP/6-31+G* method, where the optimum value of γ varies from 3.3 to 4.0. The right panel gives results for different QM methods with the cavity size contour 0.001 au, where the value of γ varies from 3.3 to 4.4. Our previous work²⁹ utilizing the SS(V)PE+FESR model with essentially the same collection of neutral and ionic solutes in water led to values of 2.9–4.0, depending on cavity size and QM method. It therefore appears that the value 3.6 is nearly optimal for various cavity sizes and QM methods in water solvent, and that value will be adopted for γ throughout the remainder of this work.

Table 1 presents optimized values for the linear parameters A , B , C , and D at each cavity contour and each QM computation method considered, with the nonlinear parameters fixed at the values of 7 au for δ and 3.6 for γ . The A parameter is not highly sensitive to the QM method or cavity size, whereas the B parameter is somewhat more so. They are of the same order of magnitude but somewhat smaller in absolute value than those previously obtained for neutral solutes in cyclohexane and benzene solvents.¹⁰ The C and D parameters, however, are quite sensitive to both the QM method and cavity size. Their values are similar to those found in our earlier study of the SS(V)PE+FESR model in water, after those were converted to au for proper comparison.

DISCUSSION OF RESULTS

Table 2 shows the mean unsigned errors from experiment categorized for the separate classes of neutral hydrocarbons, all neutrals, cations, anions, all ions, and all solutes, each with different QM methods and cavity sizes. A full listing of detailed results for each solute is given in the Supporting Information.

Many neutral hydrocarbons are hydrophobic, with positive hydration energies. Dielectric continuum models by themselves necessarily produce negative hydration energies and therefore generally perform poorly for this class of solutes. Dispersion and exchange interactions are then particularly important in this case, and furthermore aromatic hydrocarbons can accept weak hydrogen bonds from water into their π systems.^{63–65} The database utilized here contains 41 neutral hydrocarbons, with experimental hydration energies that average +0.90 kcal/mol and span a range of 7.12 kcal/mol from +2.89 to –4.23 kcal/mol. Considering different cavity sizes, the best results are obtained with the 0.001 au contour, except for HF/6-31+G*,

where the 0.0005 au contour performs slightly better. Considering different QM methods, the best results differ depending on the cavity size. Overall, the best results are obtained with the B3LYP/6-31+G* method at the 0.001 au contour, where the MUE is only 0.36 kcal/mol.

For the class of all neutral solutes the experimental hydration energies average –2.88 kcal/mol and span a range of 18.08 kcal/mol from +4.28 to –13.80 kcal/mol. Considering different cavity sizes, the best results are obtained for the 0.001 au contour with the B3LYP method and for the 0.0005 au contour with the HF method. Considering different QM methods, the best results are obtained for B3LYP/6-31+G* with the 0.0005 and 0.001 au contours, and for B3LYP/G3large with the 0.002 au contour. Overall, the best results are obtained for the B3LYP/6-31+G* method with the 0.001 au contour, where the MUE is only 0.75 kcal/mol.

For the class of all cationic solutes the experimental hydration energies average –69.2 kcal/mol and span a range of 59.4 kcal/mol from –50.9 to –110.3 kcal/mol. Considering different cavity sizes, the best results are obtained for the 0.001 au contour with all QM methods. Considering different QM methods, the best results are obtained for B3LYP/6-31+G* with the 0.0005 and 0.001 au contours, and for B3LYP/G3large with the 0.002 au contour. Overall, the best results are obtained for the B3LYP/6-31+G* method with the 0.001 au contour, where the MUE is 1.94 kcal/mol.

For the class of all anionic solutes the experimental hydration energies average –75.1 kcal/mol and span a range of 50.6 kcal/mol from –54.1 to –104.7 kcal/mol. Considering different cavity sizes, the best results are obtained for the 0.001 au contour with all QM methods. Considering different QM methods, the best results differ depending on the cavity size. Overall, the best results are obtained for the HF/6-31+G* method with the 0.001 au contour, where the MUE is 2.57 kcal/mol.

For the combined classes of all ionic solutes the experimental hydration energies average –72.4 kcal/mol and span the same range as noted above for cations. Considering different cavity sizes, the best results are obtained for the 0.001 au contour with all QM methods. Considering different QM methods, the best results differ depending on the cavity size. Overall, the best results are obtained for the HF/6-31+G* method with the 0.001 au contour, where the MUE is 2.36 kcal/mol.

For all solutes taken together the experimental hydration energies average –23.6 kcal/mol and span the range 114.6 kcal/mol from +4.28 to –110.3 kcal/mol. Considering different

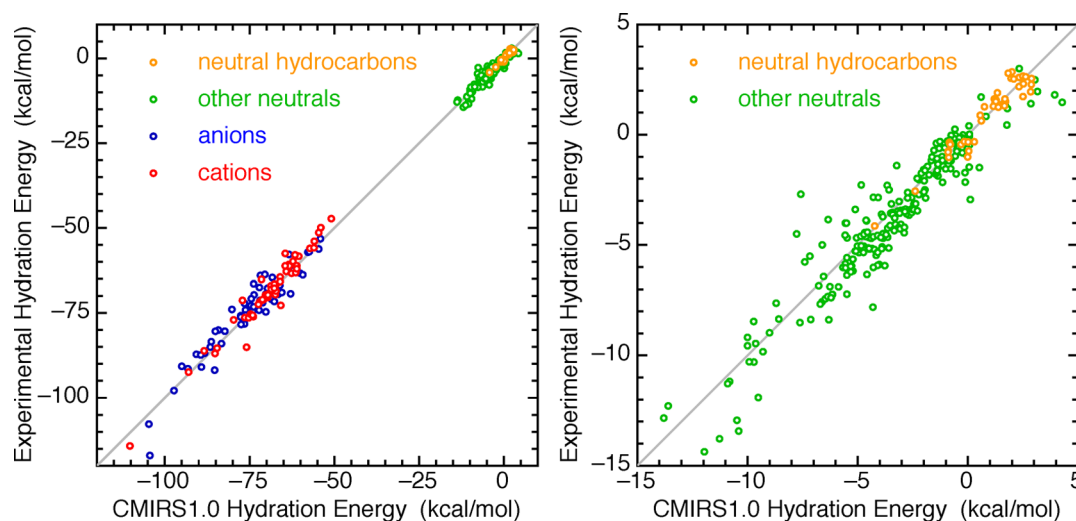


Figure 3. CMIRS1.0 hydration free energies computed with the B3LYP/6-31+G* method and the 0.001 au cavity contour vs experiment.

cavity sizes, the best results are obtained for the 0.001 au contour with all QM methods. Considering different QM methods, the best results are obtained for B3LYP/6-31+G* with the 0.0005 and 0.001 au contours, and for B3LYP/G3large with the 0.002 au contour. Overall, the best results are obtained for the B3LYP/6-31+G* method with the 0.001 au contour, where the MUE is 1.23 kcal/mol.

The present results can be compared with recent versions of the SMx series of implicit solvation models^{61,66,67} that utilized essentially the same data set of unclustered solutes while invoking many more fitting parameters. For the water solvent the present results are comparable in accuracy to the SMx MUE with various QM methods of ~ 0.2 – 0.6 kcal/mol for neutral hydrocarbons and to the SMx MUE of ~ 0.5 – 1.3 kcal/mol for all neutral solutes. SMx results have not usually been given for cations and anions separately, so comparison is not possible there. The present results are significantly better than the SMx MUE of ~ 3.1 – 5.9 kcal/mol for the combined sets of all ions and are comparable to the SMx MUE of ~ 1.3 – 2.4 kcal/mol for all solutes. It should be noted that recent versions of the SMx series of implicit solvation models for water have often utilized ions selectively clustered with a water molecule in preference to the bare ions, which has generally improved the agreement with experiment. The present results may also improve with the use of selectively clustered ions, which is a subject for possible future study.

For those few solutes where direct comparison is possible, the numerical results for dispersion and exchange contributions obtained by other groups^{11–26} are often significantly different from one another and from CMIRS1.0. This should not be too surprising, because the various approaches used to describe these effects are all quite different in nature. Although the separation of total solvation energies into individual contributions is very useful, it is phenomenological, and the various contributions cannot be distinguished experimentally.

In the interest of conciseness, the more detailed discussion of CMIRS1.0 results in the remainder of this section will consider only the B3LYP/6-31+G* method with the 0.001 au cavity size contour, which is optimal or near optimal in all cases. Very similar comparisons and conclusions pertain to the other QM methods and other contours.

Calculated attractive dispersion and repulsive exchange contributions can be individually large in magnitude in all

classes of solutes. For the respective classes of hydrocarbon, all neutral, cationic, and anionic solutes dispersion ranges up to -25.6 , -29.5 , -25.7 , and -22.8 kcal/mol and exchange ranges up to 25.7 , 32.8 , 26.4 , and 27.2 kcal/mol. However, there is substantial cancellation in all cases such that the combined dispersion plus exchange contribution only ranges from 0.2 to 4.6 kcal/mol. The average value of combined dispersion plus exchange for all solutes is 2.4 kcal/mol, which almost exactly coincides with the value 2.3 kcal/mol for the constant that was included in the SS(V)PE+FESR model²⁹ to characterize such effects on average.

Calculated FE contributions are always negative and have somewhat different ranges in the various classes of solutes. The largest (in magnitude) FE contribution among hydrocarbons is only -0.3 kcal/mol for butenyne, and the largest FE contribution among all neutrals of -4.4 kcal/mol is for the water dimer. The largest FE contribution among cations is -16.4 kcal/mol for H_3O^+ , which is a challenging case for any hydration model because of its eigen and zundel forms that are still subjects of research and discussion. The large FE result in this case is therefore probably not physically reasonable, although the computational error of 3.9 kcal/mol in the total hydration energy is not too unreasonable in this case. The largest FE contribution among anions is -37.5 kcal/mol for F^- , which is also a challenging case for any hydration model because of its very highly localized charge. The large FE contribution in this case is therefore also probably not physically reasonable, especially in view of the large computational error of -12.6 kcal/mol in the total hydration energy. Consideration of extreme cases such as these should be useful in designing future improvements to the FE model of hydrogen bonding interactions.

The present results can be compared with those previously obtained under essentially the same conditions with the simpler SS(V)PE and SS(V)PE+FESR models.²⁹ For hydrocarbons the CMIRS1.0 MUE of 0.36 kcal/mol is much better than the SS(V)PE MUE of 2.36 kcal/mol and the SS(V)PE+FESR MUE of 0.68 kcal/mol, and for all neutral solutes the CMIRS1.0 MUE of 0.75 kcal/mol is significantly better than the SS(V)PE MUE of 1.81 kcal/mol and the SS(V)PE+FESR MUE of 0.94 kcal/mol. For cations, anions, and all ions the respective CMIRS1.0 MUEs 1.94 , 2.73 , and 2.36 kcal/mol are much better than the corresponding SS(V)PE MUEs 2.95 ,

14.32, and 9.02 kcal/mol and are close to the corresponding SS(V)PE+FESR MUEs 1.80, 2.88, and 2.38 kcal/mol. Both the SS(V)PE+FESR and CMIRS1.0 results for ions are below the estimated experimental errors of 3 kcal/mol, so the small differences between these models for ions may not be meaningful. For all solutes together, the CMIRS1.0 MUE of 1.23 kcal/mol is much better than the SS(V)PE MUE of 3.98 kcal/mol and is a little better than the SS(V)PE+FESR MUE of 1.37 kcal/mol. It may be concluded that both SS(V)PE+FESR and CMIRS1.0 give improvements across the board over SS(V)PE alone and that the dispersion and exchange terms in the CMIRS1.0 model are most effective in improving over SS(V)PE+FESR the description of neutral solutes, particularly the hydrocarbons. Any improvement of CMIRS1.0 over SS(V)PE+FESR in the description of ions is not very apparent, possibly due to being masked by the larger experimental errors.

Figure 3 shows the computed CMIRS1.0 hydration energies vs experiment for all solutes, with a blow up panel for the neutral solutes. All results are seen to fall near the gray diagonal lines that would represent perfect agreement. Full numerical details for each particular solute for this and other methods and cavity contours are given in the Supporting Information.

FURTHER EXPLORATIONS

Further explorations were made in several directions to gain additional understanding and characterization of the CMIRS1.0 model. For conciseness we discuss these only for the B3LYP/6-31+G* method with the 0.001 au cavity size contour. Other QM methods and cavity contours produce similar results and conclusions.

To verify the robustness of the parametrization procedure, we carried out an exercise dividing the data set randomly into two groups of nearly equal size using one for parametrization and the other for evaluation, and then repeating this with the roles of the two groups reversed. In both cases the change in MUE was no more than 0.11 kcal/mol for any QM computational level with any cavity contour. We also considered a strategy of allowing different weights to different solutes during the parametrization, with each class weighted according to the reciprocal of its estimated experimental error, that being 0.2 kcal/mol for neutrals and 3.0 kcal/mol for ions. This again led to very similar results, the change in MUE being no more than 0.15 kcal/mol in all cases.

The dispersion and exchange contributions are correlated to a high extent, with Pearson correlation coefficients of about 0.99 in all cases. We therefore considered the possibility of omitting the expensive dispersion calculation and asserting that the combined dispersion and exchange contributions can be accommodated by change of the proportionality constant in the exchange contribution only. This leads to MUE of 0.62 kcal/mol for hydrocarbons, significantly worse than the 0.33 kcal/mol in CMIRS1.0, and to MUE = 1.55 kcal/mol for all solutes, also significantly worse than the 1.23 kcal/mol in CMIRS1.0. Reversing the idea to put the burden on a single term for dispersion without a term for exchange leads to very similar results. It is concluded that despite the high correlation between the dispersion and exchange contributions it is important to include both terms. This is because the dispersion and exchange terms cancel one another to a large extent, such that their combined effect does not correlate as well with either one separately.

Other related models of dispersion and exchange were also considered, as described above in the respective subsections

describing dispersion and exchange. In each case all parameters were reoptimized for each combination of models. The alternative dispersion model based on setting $\nabla\rho(\mathbf{r})$ to zero gave a MUE for all solutes of 1.31 kcal/mol, slightly poorer than the 1.23 kcal/mol found here. The alternative exchange model based on replacing $|\nabla\rho(\mathbf{r})|$ with $\rho(\mathbf{r})$ gave a MUE for all solutes of 1.48 kcal/mol, again poorer than found here. Simultaneous replacement of both terms with their alternative versions gave a significantly poorer MUE for all solutes of 1.63 kcal/mol. These experiments support the conclusion in our previous work¹⁰ that the alternative formulas considered for dispersion and exchange are not as effective as the ones advocated there and here.

Tests were made to determine the possible importance of an additional contribution corresponding more strictly to cavitation work. Adding either a contribution of a constant or a constant plus a term linear in the cavity volume gave negligible change in the MUE. Adding a contribution of a constant plus a term linear in the cavity surface area gave a very slight improvement in the MUE of 0.04 kcal/mol. These experiments suggest that whatever contribution cavitation work may make, it is already largely contained in the CMIRS1.0 model, probably mostly in the exchange contribution. With our philosophy of keeping the number of adjustable parameters to a minimum, we prefer not to add two more parameters to achieve such small improvement.

CONCLUSION

The value 7 au for the nonlinear dispersion damping parameter δ is nearly optimal for various cavity sizes, QM methods, and solvents, and the value 3.6 for the nonlinear field-extremum exponent γ is nearly optimal for various cavity sizes and QM methods in water. However, the linear parameters A and B governing dispersion and exchange as well the linear parameters C and D governing the field-extremum description of hydrogen bonding are all quite sensitive to those conditions and must all be separately optimized for each cavity size, QM method, and solvent to obtain good results. Among the computational approaches tested, the B3LYP/6-31+G* method with the 0.001 au cavity criterion is usually the best, or is otherwise always close to the best in performance.

Both the present CMIRS1.0 and the previous SS(V)PE+FESR models improve over the SS(V)PE dielectric continuum model alone mainly by providing a hydrogen bonding contribution that significantly improves results for neutrals and cations and dramatically improves results for anions. The dispersion and exchange models in CMIRS1.0 give additional improvement over SS(V)PE+FESR mainly for neutral solutes, especially for hydrocarbons. Little or no improvement of CMIRS1.0 over SS(V)PE+FESR is found for ions, possibly because of being masked by the larger experimental errors there.

No evidence is found for a significant contribution from cavitation beyond whatever part of it is already subsumed in the CMIRS1.0 model.

The CMIRS1.0 model used in conjunction with an isodensity criterion for the cavity provides a good account of hydration energies even while maintaining the simplifying fiction of a structureless water solvent. It uses only seven adjustable parameters. One parameter is the value of the isodensity contour that governs the solute size and the other six parameters are associated with the DEFESR contributions. For example, for solutes in the database used CMIRS1.0 gives

MUE as low as 0.8 kcal/mol for neutrals, where the experimental results span a range of 18.1 kcal/mol, and as low as 2.4 kcal/mol for ions, where the experimental results span a range of 69.2 kcal/mol. Compared to recent SMx models that invoke many more adjustable parameters, CMIRS1.0 gives comparable performance on neutrals and significantly better performance on ions.

The success of CMIRS1.0 with so few parameters implies that essentially correct physics is being captured in each of the contributing terms. By virtue of formulating each contribution as a functional of the calculated solute charge density, the various interactions are able to respond to any particular vagaries of the solute electronic structure. Those observations, together with the fact that the database used contains solutes with many different functional groups, lead to an expectation that the model will also perform well for many other solutes in water. The present results are also encouraging that this initial version of the model constitutes a valuable framework for development of future refinements to further increase its accuracy.

■ ASSOCIATED CONTENT

■ Supporting Information

Tables listing hydration energies calculated for all solutes with various QM methods and various cavity sizes. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Present Address

[†]Inorganic Chemistry Group, Department of Chemistry, St. Petersburg State University, University Pr. 26, Old Peterhof, 198504, Russia.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

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-4984 from the Notre Dame Radiation Laboratory.

■ REFERENCES

- (1) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094.
- (2) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161–2200.
- (3) Luque, F.; Curutchet, C.; Munoz-Muriedas, J.; Bidon-Chanal, A.; Soteras, I.; Morreale, A.; Gelpi, J.; Orozco, M. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3827–3836.
- (4) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.
- (5) Mennucci, B. *Comp. Mol. Science* **2012**, *2*, 386–404.
- (6) Böttcher, C. J. F. *Theory of Electric Polarization*; Elsevier: Amsterdam, 1973; p 42ff.
- (7) Miertuš, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117–129.
- (8) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (9) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305–8311.
- (10) Pomogaeva, A.; Chipman, D. M. *J. Phys. Chem. A* **2013**, *117*, 5812–5820.
- (11) Linder, B. *J. Chem. Phys.* **1960**, *33*, 668–675.
- (12) Linder, B. *J. Chem. Phys.* **1962**, *37*, 963–966.
- (13) Costa Cabral, B. J.; Rinaldi, D.; Rivail, J.-L. *C. R. Acad. Sci. Paris, Ser. II* **1984**, *298*, 675–678.
- (14) Rinaldi, D.; Costa Cabral, B. J.; Rivail, J.-L. *Chem. Phys. Lett.* **1986**, *125*, 495–499.
- (15) Aguilar, M. A.; Olivares del Valle, F. J. *Chem. Phys.* **1989**, *138*, 327–336.
- (16) Olivares del Valle, F. J.; Aguilar, M. A. *J. Mol. Struct. (THEOCHEM)* **1993**, *99*, 25–47.
- (17) Rösch, N.; Zerner, M. C. *J. Phys. Chem.* **1994**, *98*, 5817–5823.
- (18) Amovilli, C. *Chem. Phys. Lett.* **1994**, *229*, 244–249.
- (19) Amovilli, C.; Mennucci, B. *J. Phys. Chem. B* **1997**, *101*, 1051–1057.
- (20) Weijo, V.; Mennucci, B.; Frediani, L. *J. Chem. Theory Comput.* **2010**, *6*, 3358–3364.
- (21) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616–627.
- (22) Floris, F. M.; Tomasi, J.; Pascual-Ahuir, J. L. *J. Comput. Chem.* **1991**, *12*, 784–791.
- (23) Floris, F. M.; Tani, A.; Tomasi, J. *Chem. Phys.* **1993**, *169*, 11–20.
- (24) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 3649–3659.
- (25) Duignan, T. T.; Parsons, D. F.; Ninham, B. W. *J. Phys. Chem. B* **2013**, *117*, 9412–9420.
- (26) Duignan, T. T.; Parsons, D. F.; Ninham, B. W. *J. Phys. Chem. B* **2013**, *117*, 9421–9429.
- (27) Klamt, A. *COSMO-RS. From Quantum chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier: Amsterdam, 2005; p 55.
- (28) Pomogaeva, A.; Thompson, D. W.; Chipman, D. M. *Chem. Phys. Lett.* **2011**, *511*, 161–165.
- (29) Pomogaeva, A.; Chipman, D. M. *J. Chem. Theory Comput.* **2011**, *7*, 3952–3960.
- (30) Chipman, D. M. *J. Chem. Phys.* **1997**, *106*, 10194–10206.
- (31) Chipman, D. M. *J. Chem. Phys.* **2000**, *112*, 5558–5565.
- (32) Chipman, D. M. *J. Chem. Phys.* **2002**, *116*, 10129–10138.
- (33) Chipman, D. M. *Theor. Chem. Acc.* **2002**, *107*, 80–89.
- (34) Zhan, C.-G.; Bentley, J.; Chipman, D. M. *J. Chem. Phys.* **1998**, *108*, 177–192.
- (35) Chipman, D. M. *J. Chem. Phys.* **2006**, *124*, 224111.
- (36) Vilks, M. J.; Zhan, C.-G. *J. Chem. Phys.* **2008**, *129*, 194109.
- (37) Amovilli, C.; Filippi, C.; Floris, F. M. *J. Chem. Phys.* **2008**, *129*, 244106.
- (38) Cancès, E.; Mennucci, B. *J. Chem. Phys.* **2001**, *114*, 4744–4745.
- (39) Mennucci, B.; Cammi, R.; Tomasi, J. *J. Chem. Phys.* **1998**, *109*, 2798–2807.
- (40) Cossi, M.; Barone, V. *J. Phys. Chem. A* **2000**, *104*, 10614–10622.
- (41) Vydrov, O. A.; Van Voorhis, T. *Phys. Rev. A* **2010**, *81*, 062708:1–6.
- (42) Vydrov, O. A.; Van Voorhis, T. *Phys. Rev. Lett.* **2009**, *103*, 063004:1–4.
- (43) Andersson, Y.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **1996**, *76*, 102–105.
- (44) Gor'kov, L. P.; Pitaevskii, L. P. *Sov. Phys. Dokl.* **1964**, *8*, 788–790.
- (45) Herring, C.; Flicker, M. *Phys. Rev. A* **1964**, *134*, A362–A366.
- (46) Tang, K. T.; Toennies, J. P.; Yiu, C. L. *Int. Rev. Phys. Chem.* **1998**, *17*, 363–406.
- (47) Sadlej, J.; Buch, V.; Kazimirski, J. K.; Buck, U. *J. Phys. Chem. A* **1999**, *103*, 4933–4947.
- (48) Chipman, D. M. *J. Chem. Phys.* **2003**, *118*, 9937–9942.
- (49) Chipman, D. M.; Chen, F. W. *J. Chem. Phys.* **2006**, *124*, 144507:1–5.
- (50) 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, 2009.
- (51) Ben-Naim, A. *J. Phys. Chem.* **1978**, *82*, 792–803.
- (52) 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*, 7787–7794.
- (53) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

- (54) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (55) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (56) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (57) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. *J. Comput. Chem.* **1983**, *4*, 294–301.
- (58) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764–7776.
- (59) Zhan, C.-G.; Chipman, D. M. *J. Chem. Phys.* **1998**, *109*, 10543–10558.
- (60) Zhan, C.-G.; Chipman, D. M. *J. Chem. Phys.* **1999**, *110*, 1611–1622.
- (61) 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*, 1109–1117.
- (62) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (63) Atwood, J. L.; Hamada, F.; Robinson, K. D.; Orr, G. W.; Vincent, R. L. *Nature* **1991**, *349*, 683–684.
- (64) Gotch, A. J.; Zwier, T. S. *J. Chem. Phys.* **1992**, *96*, 3388–3401.
- (65) Suzuki, S.; Green, P. G.; Bumgarner, R. E.; Dasgupta, S.; Goddard, W. A.; Blake, G. A. *Science* **1992**, *257*, 942–944.
- (66) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- (67) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 609–620.