JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

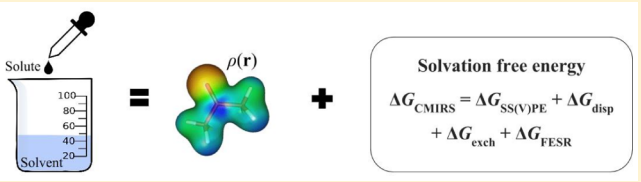
Reparameterization of an Accurate, Few-Parameter Implicit Solvation Model for Quantum Chemistry: Composite Method for Implicit Representation of Solvent, CMIRS v. 1.1

Zhi-Qiang You and John M. Herbert*

Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States

Supporting Information

ABSTRACT: CMIRS (composite method for implicit representation of solvent) is a relatively new implicit solvation model that adds terms representing solute−solvent dispersion, Pauli repulsion, and hydrogen bonding to a continuum treatment of electrostatics. A small error in the original implementation of the dispersion term, but one that can modify dispersion energies by up to 8 kcal/mol in some cases, necessitates refitting the parameters in the model, which we do here. We refer to the modified implementation and parameter set as CMIRS v. 1.1. While the dispersion energies change in nontrivial ways, an increase in the attractive dispersion term in the new implementation is largely offset by an increase in the Pauli repulsion during the fitting process, such that overall statistical errors are virtually unchanged with respect to v. 1.0 of the model, for a large database of experimental solvation free energies for molecules and ions. Overall, we obtain mean unsigned errors of <0.7 kcal/mol when the solvent is cyclohexane or benzene, <1.5 kcal/mol for water, and <2.8 kcal/mol for dimethyl sulfoxide and acetonitrile, despite using no more than five empirical parameters per solvent. For the important but difficult case of ionic solutes in water, mean unsigned errors are <2.9 kcal/mol.



I. INTRODUCTION

Implicit solvation modeling is a popular and low-cost way to estimate free energies of solvation,[1−5] which in the context of a quantum-mechanical (QM) description of the solute (the exclusive focus of this work) has historically meant at the very least a solution of Poisson's equation or its equivalent for the electrostatic energy associated with a continuum dielectric description of the solvent.[6−9] This requires a suitable partition of the system into an atomistic part (the solute) and a continuum part (the solvent), and this partition constitutes the "solute cavity;" see Figure 1. The solute−continuum interaction can then be interpreted as the free energy of solvation (ΔG), and sampling over the configuration space of explicit solvent molecules is not required.

That said, whereas continuum electrostatics methods such as Poisson's equation or else the polarizable continuum model[3,9,10] (PCM) can account for long-range solute−solvent interactions, an *accurate* model for solvation free energies must also include a treatment of short-range, nonelectrostatic interactions.[4,11] Various models decompose these interactions in different ways, but usually the nonelectrostatic terms attempt to model all or most of the following:[3,4,11,12] solute−solvent dispersion (van der Waals) interactions, Pauli ("exchange") repulsion between solute and solvent, the work associated with forming the solute cavity within the dielectric medium (the so-called "cavitation energy"), hydrogen-bonding and other "specific" interactions due to the molecular structure of the solvent, and changes in the structure (and therefore the entropy) of the neat solvent upon introduction of the solute. A variety of approaches for introducing these nonelectrostatic interactions have been used in conjunction with a continuum treatment of electrostatics.[4,5,12−22]

Recently, Pomogaeva and Chipman[22−24] introduced a new implicit solvent model that they call the "composite method for implicit representation of solvent" (CMIRS). This model consists of a self-consistent treatment of solute−continuum electrostatics, wherein the solute is described quantum-mechanically and the interaction with the continuum is described using the "surface and volume polarization for electrostatics" [SS(V)PE] method,[8] which is closely related[9,10,25,26] to the "integral equation formulation" of the PCM (IEF-PCM).[6] To this electrostatics calculation, the CMIRS approach adds a solute−solvent dispersion term that is modeled upon the nonlocal "VV09" van der Waals dispersion density functional of Vydrov and Van Voorhis,[27] a Pauli repulsion contribution that depends upon the tail of the solute's electron density that extends beyond the solute cavity, and a hydrogen-bonding correction based on the maximum and minimum values of the normal component of the electric field generated by the solute at the cavity surface. Although this model must be independently parametrized for each solvent of interest, it uses no more than five empirical parameters per solvent, with parameters for water,[23] acetonitrile (CH_3CN),[24] dimethyl sulfoxide (DMSO),[24] benzene,[22] and cyclohexane[22] having been reported thus far.

For these solvents, the accuracy of CMIRS versus experimental solvation free energies ΔG is at least as good as

Received: June 26, 2016
Published: July 29, 2016

DOI: 10.1021/acs.jctc.6b00644
J. Chem. Theory Comput. 2016, 12, 4338−4346

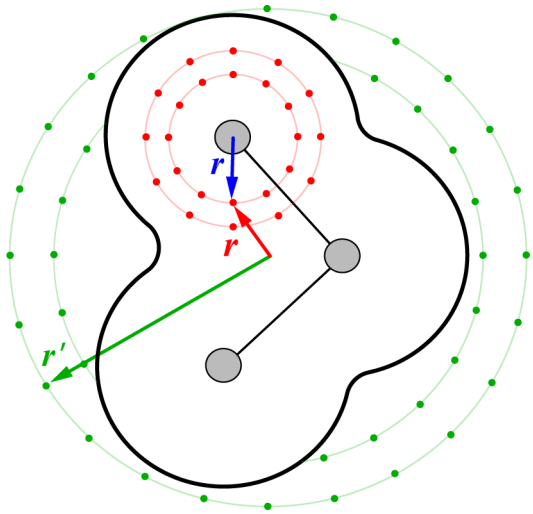


Figure 1. Schematic description of the vectors **r** and **r**′ in eqs 3 and 6, for a triatomic solute. Red points represent a few of the atom-centered Lebedev quadrature points used in the DFT calculation, which are also used to integrate the solute's charge density for CMIRS. Green points lie on the single-center set of concentric Lebedev grids that is used to determine the isodensity surface (heavy black curve) and to integrate the solvent region that lies outside of this surface. The blue vector shows how **r** was erroneously defined in previous work,[22−24] whereas the red vector **r** is the definition used here, so that **r** and **r**′ share a common origin coinciding with the center of the Lebedev grids used to determine the isodensity surface.

that of the popular "SM*x*" solvation models developed by Truhlar and co-workers,[4,12,28,29] while requiring fewer fitting parameters. In fairness, it should be pointed out that later versions of SM*x* (specifically SM8,[28] SM12,[29] and SMD[12]) are intended to be "universal" solvation models that are available for a very large number of solvents with a single set of parameters,[4] whereas CMIRS is currently parametrized only for the five solvents mentioned above, and the parameters are solvent-dependent. Nevertheless, the outstanding performance of CMIRS is impressive, given its limited parameter set.

The original implementation of the model,[22−24] CMIRS v. 1.0, is available in the GAMESS quantum chemistry program.[30] In the course of implementing this model in the Q-Chem program,[31] we discovered an error in the original implementation of the solute−continuum dispersion interaction, correction of which changes the dispersion energies by as much as 7.8 kcal/mol for some molecules in the data set used to parametrize the model. As such, correct numerical implementation of the dispersion formula originally suggested in ref 22 requires refitting the parameters of the model and reassessing its accuracy. That is the topic of the present work, and we refer to the new model as CMIRS v. 1.1.

II. THEORY

II.A. CMIRS. The CMIRS models the Gibbs free energy of solvation, ΔG, as[22−24]

$$\Delta G_{\mathrm{CMIRS}} = \Delta G_{\mathrm{SS(V)PE}} + \Delta G_{\mathrm{DEFESR}} \tag{1}$$

where $\Delta G_{\mathrm{SS(V)PE}}$ is the continuum electrostatics contribution, which includes an approximate but accurate description of the "volume polarization" that arises from the tail of the solute wave function, which may penetrate beyond the solute cavity.[8] The SS(V)PE model defines the solute cavity to be an isocontour of the solute's charge density, and as in the original CMIRS model,[22−24] we determine this isodensity cavity using a single-center approach[32,33] and discretize it with a single-center Lebedev grid.[34] The single-center approach is highly efficient but can fail for certain nonspherical solute geometries.[34] The second term in eq 1 contains the short-range dispersion, exchange, and "field-extremum short-range" (DEFESR) interactions:

$$\Delta G_{\mathrm{DEFESR}} = \Delta G_{\mathrm{disp}} + \Delta G_{\mathrm{exch}} + \Delta G_{\mathrm{FESR}} \tag{2}$$

Whereas the SS(V)PE electrostatic energy is updated at each self-consistent field iteration, and thus determined self-consistently alongside the solute's charge density, the DEFESR components are evaluated only once, using the converged charge density.

The dispersion contribution to eq 2 is adapted from the VV09 density functional,[27] a nonlocal correlation functional designed to account for the van der Waals interaction between atoms and molecules in the gas phase. In the present context, the working formula for the solute−solvent dispersion interaction is

$$\Delta G_{\mathrm{disp}} = A \int_{\mathrm{solute}} \mathrm{d}^3\mathbf{r}\, S(\mathbf{r};\, \bar{\rho}_{\mathrm{solvent}})\, I(\mathbf{r};\, \delta) \tag{3}$$

with

$$S(\mathbf{r};\, \bar{\rho}_{\mathrm{solvent}}) = \frac{\rho(\mathbf{r})}{w[\rho(\mathbf{r})](w[\rho(\mathbf{r})] + \bar{\rho}_{\mathrm{solvent}}^{1/2})} \tag{4}$$

$$w[\rho(\mathbf{r})] = \left(\rho(\mathbf{r}) + \frac{3c}{4\pi}\left|\frac{\hat{\nabla}\rho(\mathbf{r})}{\rho(\mathbf{r})}\right|^4\right)^{1/2} \tag{5}$$

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

(All formulas are given in atomic units, where $\hbar$, e, m_e, and $1/4\pi\epsilon_0$ each have the value of unity.) The quantity $\bar{\rho}_{\mathrm{solvent}}$ represents the solvent's average electron density, as taken from experiment. The constant $c = 0.0089$ is taken unchanged from the value that was optimized to fit a test set of gas-phase intermolecular interaction energies in the development of VV09.[27] The linear parameter A in eq 3 and the nonlinear damping parameter δ in eq 6 are to be determined empirically in order to reproduce experimental solvation energies. A linear factor of $\bar{\rho}_{\mathrm{solvent}}^{1/2}$ that would otherwise appear in eq 3 has been absorbed into the A parameter for convenience, and the value of A should be negative because the dispersion interaction is attractive. Note that evaluation of $I(\mathbf{r};\delta)$ in eq 6 requires integration over the solvent (defined as those Lebedev points exterior to the solute cavity) and that to evaluate the integral in eq 3 this integration over solvent must be done for each solute grid point, that is, for each value of **r** for which $\rho(\mathbf{r})$ is not negligible. In practice, this means that integration over **r** extends somewhat beyond the solute cavity.

The tail of $\rho(\mathbf{r})$ that penetrates beyond the solute cavity also contributes to the exchange interaction with the solvent, ΔG_{exch}, which in CMIRS 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.[35,36] This formalism suggests the functional form

DOI: 10.1021/acs.jctc.6b00644
J. Chem. Theory Comput. 2016, 12, 4338−4346

$$\Delta G_{\text{exch}} = B \int_{\text{solvent}} d^3\mathbf{r} |\hat{\nabla}\rho(\mathbf{r})| \quad (7)$$

where B is an empirical parameter. Note that "exchange" in this context denotes the QM antisymmetry requirement on the total (hypothetical solute + solvent) wave function, including the *intra*-molecular exchange effect that is always negative and the Pauli repulsion (penetration) effect that is always positive.[37] In the region of van der Waals interactions, the penetration term is normally about twice as large as the stabilizing intramolecular exchange term;[14] therefore the value of B should be positive and ΔG_{exch} reduces the solvation energy.

The final component of CMIRS is a field-effect short-range (FESR) term that is intended to describe specific hydrogen-bonding interactions.[20] Hydrogen bond strength in water, as estimated by the distribution of O−H vibrational frequency shifts, correlates well with local electric field strength in classical molecular dynamics simulations,[38,39] so it makes some sense to use the normal component of the electric field at the cavity surface as an indicator of hydrogen-bond affinity, since this quantity is computed anyway in SS(V)PE calculations. In particular, the FESR model uses the minimum and maximum values (F_{min} and F_{max}) of the outgoing normal electric field produced by the solute, evaluated at the cavity surface. These values are assumed to describe donation or acceptance of a hydrogen bond, respectively, by the solvent. If the outgoing normal field is strictly positive, then F_{min} is set to zero, and if it is strictly negative, then F_{max} is set to zero. The FESR contribution to ΔG is parametrized in the form[20]

$$\Delta G_{\text{FESR}} = C|F_{\text{min}}|^{\gamma} + DF_{\text{max}}^{\gamma} \quad (8)$$

where the linear proportionality constants C and D as well as the exponent γ are empirical parameters. (The exponent γ is present because there is no reason to believe that hydrogen bond strength should correlate *linearly* with field strength.) As noted by Pomogaeva and Chipman,[20,23,24] a serious limitation of the FESR model is that it describes, at best, only the single strongest donor and single strongest acceptor hydrogen bond sites on the solute.

Note that in the DEFESR model of eq 2, there is no explicit cavitation contribution to the solvation free energy. As reported by Pomogaeva and Chipman,[22−24] an explicit treatment of cavitation (e.g., an energy penalty depending on the volume of the solute cavity) afforded only a very slight statistical improvement in solvation energies as compared to the experiment, at the cost of an increased number of empirical parameters. Those authors suggest that the cavitation work may already be contained implicitly in CMIRS, probably mostly in the exchange term.[23] Indeed, the exchange component has comparable extent to the repulsion term plus the cavitation terms in the formalism of Amovilli and Mennucci.[14] On the other hand, the lack of an explicit cavitation term or the ability to describe multiple hydrogen bond donor and/or acceptor sites suggests that CMIRS may be most appropriate for modeling molecules that are about the same size as those in the training set, which currently ranges in size up to decamethyltetrasiloxane (48 atoms), *n*-pentadecane (48 atoms), *n*-hexadecane (51 atoms), and ethyloctadecanoate (63 atoms).

II.B. New Implementation. The original implementation[22−24] of CMIRS contained an error in the evaluation of the integral in eq 6, the nature of which can be understood from Figure 1, which provides a schematic view of how the integrals over $\mathbf{r}$ (in eq 3) and $\mathbf{r}'$ (in eq 6) are evaluated. The former uses the same atom-centered quadrature grids that are used to integrate the exchange-correlation functional in a DFT calculation, whereas the latter uses the concentric single-center Lebedev grids that are used to determine the isodensity cavity surface. Mathematically, both $\mathbf{r}$ and $\mathbf{r}'$ should share a common coordinate origin, which we take to be the center of the isodensity Lebedev grids. In the original implementation of CMIRS, however, the origin for $\mathbf{r}$ was incorrectly taken to be whatever atom was associated with the DFT grid point in question. This is the blue vector in Figure 1, whereas in the present work the red vector is used instead to define $\mathbf{r}$, so that the origin is the same for all grid points.

We have verified in our implementation that if we do indeed use the blue vector in Figure 1 to define $\mathbf{r}$, then we are able to reproduce the dispersion energies provided in the Supporting Information to ref 24, to many significant digits. All other contributions to the CMIRS solvation energy are identical in both implementations, but the significant change in the dispersion energies necessitates reparameterization of the model.

II.C. Computational Details. Solute molecules reported in refs 22−24 were tested, with optimized gas-phase geometries obtained from the Minnesota Solvation Database, v. 2009,[40] and not further optimized. Isodensity cavity surfaces were determined as described in ref 34 using single-center Lebedev grids with 1202 angular grid points. Although this single-center approach can in some cases fail to generate an isodensity cavity surface, it is highly efficient in cases where it succeeds because the integration over solvent in eq 6 can be separated into radial and angular parts, the former of which can be integrated analytically.[22]

The SS(V)PE treatment of electrostatics contains a single empirical parameter, namely, the isodensity value ρ_0, and we test values $\rho_0 = 0.0005$ and 0.0010 au that were found to afford good results in previous work.[8,22−24,33] Note that the SS(V)PE solvation model uses the symmetric form of the matrix $\mathbf{K}$ in the PCM equation $\mathbf{Kq} = \mathbf{Rv}$ (in the notation of refs 9 and 26), as distinguished from the asymmetric form used in IEF-PCM. Dielectric constants and values of $\overline{\rho}_{\text{solvent}}$ were taken from refs 22−24 and can also be found in the Supporting Information of the present work. All calculations were performed at the B3LYP/6-31+G* level for the solute, using a locally modified version of Q-Chem.[31]

Empirical parameters in CMIRS were optimized to best reproduce experimental solvation energies in the Minnesota Solvation Database, v. 2009.[40] Specifically, for a fixed value of γ, we optimize the linear parameters A, B, C, and D in the DEFESR model using the Nelder−Mead algorithm from the Python SciPy package[41] in order to minimize either the mean unsigned error (MUE)

$$\text{MUE}(\gamma) = \min_{A,B,C,D} \left(\frac{1}{N} \sum_{i=1}^{N} |\Delta G_{\text{expt}} - \Delta G_{\text{CMIRS}}(A, B, C, D; \gamma)| \right) \quad (9)$$

or else the root meant square error (RMSE)

DOI: 10.1021/acs.jctc.6b00644
J. Chem. Theory Comput. 2016, 12, 4338−4346

Table 1. Optimized[a] CMIRS Parameters[b] (in a.u.) for Various Solvents

		ρ_0 = 0.0005 au					ρ_0 = 0.0010 au				
solvent	reference	A	B	C	D	γ	A	B	C	D	γ
benzene	ref 22	−0.01016	0.06533				−0.01154	0.03442			
	this work	−0.00572	0.01116				−0.00522	0.01294			
cyclohexane[c]	ref 22	−0.01730	0.06039				−0.01496	0.04380			
	this work	−0.00721	0.05618				−0.00938	0.03184			
DMSO[d]	ref 24	−0.003758	0.017585		−1041.53	4.4	−0.014401	0.059515		−253.58	4.3
	this work	−0.002523	0.011757		−817.93	4.3	−0.009510	0.044791		−162.07	4.1
DMSO[e]	ref 24	**0. 001392**	**−0. 025825**		−2989.37	5.0	−0.009872	0.036192		−696.65	5.0
	this work	**0. 000728**	**−0. 022288**		−12939.70	5.6	−0.006330	0.025298		−2639.89	5.6
CH_3CN[d]	ref 24	−0.000463	0.006345		−0.43405	1.2	−0.004984	0.028574		−0.32009	1.3
	this work	−0.003805	0.032230		−0.44492	1.2	−0.008178	0.045278		−0.33914	1.3
CH_3CN[e]	ref 24	**0. 003232**	**−0. 035571**		−1.48289	2.0	−0.001832	0.006349		−1.54873	2.5
	this work	**0. 001436**	**−0. 024959**		−1.48231	2.0	−0.002555	0.010156		−2.36788	2.7
water	ref 23	−0.009663	0.063101	−1840.0	−70.873	3.6	−0.010825	0.045576	−944.4	−17.817	3.6
	this work	−0.006496	0.050833	−566.7	−30.503	3.2	−0.006736	0.032698	−1249.6	−21.405	3.7

[a]The damping parameter δ = 7 au is fixed, not optimized. [b]Unphysical values of the parameters A and B are highlighted in bold. [c]In ref 22, the damping parameter δ = 6 au was used. [d]Experimental solvation energies for ions were obtained based on the proton solvation free energy from Kelly et al.[43] [e]Experimental solvation energies for ions were obtained based on the proton solvation free energy from Fawcett.[44]

Table 2. Mean Unsigned Errors in CMIRS v. 1.1 Solvation Energies (in kcal/mol), As Compared to Experimental Values

	ρ_0 = 0.0005 au						ρ_0 = 0.0010 au					
solvent	all solutes	hydro-carbons	all neutrals	cations	anions	all ions	all solutes	hydro-carbons	all neutrals	cations	anions	all ions
benzene	0.72						0.64					
cyclohexane	0.38						0.43					
DMSO[a]	2.32		0.52	1.00	2.58	2.49	2.46		2.93	0.68	2.52	2.41
DMSO[b]	—[c]						2.24		0.91	0.45	2.49	2.37
CH_3CN[a]	2.62		2.31	2.64	2.67	2.65	2.80		2.89	2.84	2.73	2.80
CH_3CN[b]	—[c]						2.08		1.30	2.25	2.04	2.16
water	1.53	0.59	0.93	2.86	3.01	2.94	1.25	0.43	0.78	1.84	2.82	2.36

[a]Experimental solvation energies for ions were obtained based on the proton solvation free energy from Kelly et al.[43] [b]Experimental solvation energies for ions were obtained based on the proton solvation free energy from Fawcett.[44] [c]Excluded, because either A or B acquired an unphysical sign during the fitting process.

$$\mathrm{RMSE}(\gamma) = \min_{A,B,C,D}\left[\frac{1}{N}\sum_{i=1}^{N}\left|\Delta G_{\mathrm{expt}} - \Delta G_{\mathrm{CMIRS}}(A, B, C, D; \gamma)\right|^2\right]^{1/2} \quad (10)$$

We then minimize the one-dimensional function MUE(γ) or RMSE(γ) according to the MUE criterion, in order to obtain the final parameter, γ. The recommended values of the linear parameters are fit to minimize the RMSE criterion in eq 10, for reasons discussed below, whereas the robustness tests in Section III.B use the MUE criterion. In cases where both procedures are well-behaved, results are about the same.

Throughout this work, the damping parameter in eq 6 is fixed at δ = 7 au (about 3.7 Å), a value that was optimized previously.[22−24] This value is about twice the van der Waals radius for most atoms in the data set, indicating that dispersion is considered only at intermolecular distances larger than the van der Waals contact distance. Values of ΔG_{disp} predicted by eq 3 are within the typical range of values determined in studies of gas-phase intermolecular dispersion energies.[42]

III. RESULTS AND DISCUSSION

CMIRS parameters optimized in this work, using the RMSE criterion in eq 10, are listed in Table 1 alongside those reported previously for v. 1.0 of the model.[22−24] (While we use the RMSE criterion of eq 10 to optimize the parameters throughout this work, to facilitate easy comparison with other implicit solvation models we report error statistics in terms of MUEs, which are evaluated at the RMSE-optimized parameters.) Although the dispersion energies change significantly as compared to the implementation in v. 1.0, at least in cases where ΔG_{disp} is large, the overall MUEs actually change very little as compared to v. 1.0. For example, in cyclohexane, we obtain a MUE of 0.43 kcal/mol (CMIRS v. 1.1), as compared to the value of 0.41 kcal/mol (CMIRS v. 1.0) that was reported previously,[22] despite changes averaging 4.5 kcal/mol in ΔG_{disp}. As such, we will highlight only the most significant changes. Unless stated otherwise, the following discussion applies equally well to results obtained using either ρ_0 = 0.0005 au or ρ_0 = 0.0010 au to define the isodensity cavity.

III.A. Optimal Parameters and Mean Unsigned Errors.

For the two nonpolar solvents, cyclohexane and benzene, we do not include the hydrogen-bonding correction ΔG_{FESR}, and thus no values are reported in Table 1 for the parameters C, D, and γ in these cases. The dispersion parameters A for these two solvents are reduced by a factor of 2 as compared to those reported previously;[22] the exchange parameters B are reduced significantly only for benzene. These changes reflect that the actual values of ΔG_{disp} change significantly because values of the integral in eq 3 are simply different in our implementation. We

DOI: 10.1021/acs.jctc.6b00644
J. Chem. Theory Comput. 2016, 12, 4338−4346

find, however, that upon refitting of the parameters with this new implementation of ΔG_{disp}, the exchange energies change significantly as well. For example, in the case of cyclohexane using $\rho_0 = 0.001$ au, dispersion energies change by an average of 4.5 kcal/mol for the 87 molecules in the data set, relative to those obtained using v. 1.0 of the model, but at the same time the average change in the exchange energies is −4.5 kcal/mol! We will return to this observation below, as significant cancellation between dispersion and exchange appears to be a feature of the model.

Using the new parameters, we obtain MUEs of no larger than 0.72 kcal/mol in benzene and 0.43 kcal/mol in cyclohexane. (The precise MUE depends on the choice of ρ_0.) MUEs for v. 1.1 of the model, for all solvents, are reported in Table 2. Figure 2 plots the CMIRS v. 1.1 solvation energies versus experiment

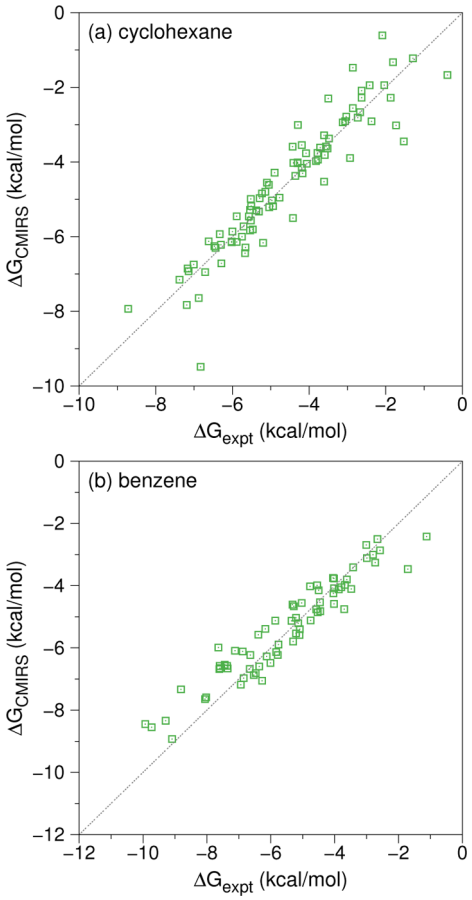


Figure 2. CMIRS v. 1.1 solvation energies in (a) benzene and (b) cyclohexane, versus experimental values. The CMIRS calculations use an isodensity contour $\rho_0 = 0.001$ au, and the solute is described at the B3LYP/6-31+G* level.

for $\rho_0 = 0.001$ au, and we see that the data are clustered tightly along the diagonal that represents perfect agreement with the experiment.

Moving on to DMSO and CH_3CN, we set $C = 0$ by fiat (as in CMIRS v. 1.0[24]), since neither of these solvents is expected to serve as hydrogen bond donor. In discussing the data for polar solvents, where some single-ion solvation free energies are included in the data set, it is important to bear in mind that extracting single-ion solvation energies from experiments on ion pairs requires an absolute reference, which is usually taken to be the solvation free energy of the proton. However, two different values for this reference in DMSO and CH_3CN can be found in the recent literature,[43,44] so there are really two different sets of experimental data for these two solvents, depending on which reference is used. For all of the polar solvents, we report parameters that have been optimized separately depending on whether the experimental data for ions are tied to Fawcett's value[44] of the proton solvation energy or else the value reported by Kelly et al.[43] (The latter is a revised version, using additional data, of the value originally derived by Coe and co-workers,[45] as discussed also in ref 46.) Careful inspection of Table 1 reveals two cases (specifically DMSO and CH_3CN with $\rho_0 = 0.0005$ au, using Fawcett's value) where the fitting procedure yields values of A and B whose signs are unphysical, suggesting repulsive dispersion but attractive exchange. (These cases are highlighted in bold in Table 1.) Pomogaeva and Chipman previously reported such an anomaly, in precisely the same two cases.[24]

Before discussing the optimal parameters for DMSO and CH_3CN, we demonstrate why we choose eq 10 rather than eq 9 for parameter optimization. Figure 3a plots the MUE for DMSO as a function of the nonlinear parameter γ, but where in one case the linear parameters were obtained by minimizing the RMSE at fixed γ, whereas in the other case they are optimized to minimize the MUE. In the latter case, the MUE oscillates dramatically as a function of γ, for $\gamma > 7$ and $\gamma < 3$. When the

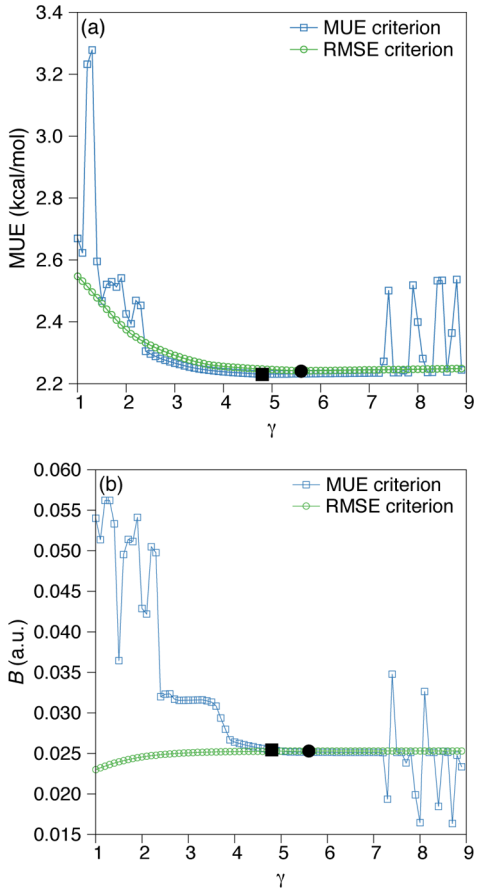


Figure 3. (a) MUEs and (b) B parameters for DMSO, obtained by minimizing the linear parameters at fixed γ according to either the MUE (eq 9) or the RMSE criterion (eq 10). The black symbols indicate the optimal values. Calculations were performed using $\rho_0 = 0.001$ au, B3LYP/6-31+G*, and Fawcett's value[44] of the proton solvation energy.

DOI: 10.1021/acs.jctc.6b00644
J. Chem. Theory Comput. 2016, 12, 4338−4346

linear parameters are optimized according to the RMSE criterion, the MUE remains well-behaved even in these cases, suggesting that the RMSE is the better objective function. Figure 3b shows that the optimal value of the B parameter (at fixed γ) is also wildly oscillatory when the MUE criterion is employed but is well-behaved when optimized (simultaneously with the other linear parameters) to minimize the RMSE.

Excluding the two anomalous cases with unphysical optimized parameters, we can say in general that our v. 1.1 values of A and B are smaller in magnitude for DMSO but larger for CH_3CN, as compared to the values reported for v. 1.0.[24] Optimal values for γ and D in ΔG_{FESR} are quite comparable to those obtained in previous work, except for the case of DMSO based on experimental data tied to Fawcett's proton reference, where D is dramatically different, and γ is more different than what is seen for other solvents (see Table 1). Recall that $\Delta G_{FESR} = DF^{\gamma}_{max}$ for these two solvents, so it is not altogether surprising that a change from $\gamma = 5.0$ (ref 24) to $\gamma = 5.6$ (this work) is compensated by a large change in the corresponding coefficient, D. The relatively large change in γ occurs because the error profile as a function of γ is quite flat for $4 < \gamma < 9$; see Figure 3a. Unfortunately, the data set used to fit the parameters for DMSO consists overwhelmingly of anions, with very few neutral or cationic solutes, and since small anions are unlikely to form hydrogen bonds with DMSO, the number of data points that contribute significantly to the optimization of D and γ is quite limited in this case; see Figure 4a. The situation is less severe for CH_3CN, as shown in Figure 4b. We have been unable to locate significant additional experimental data in DMSO that might resolve this situation.

Figure 5 plots the calculated CMIRS v. 1.1 solvation free energies versus the Fawcett-based experimental values, using $\rho_0 = 0.001$ au. The MUE is 2.24 kcal/mol for DMSO and 2.08 kcal/mol for CH_3CN, across all solutes, and only slightly greater for the ions-only subset (see Table 2), which likely reflects the fact that most of the experimental data points for these two solvents are in fact ions. MUEs remain <3 kcal/mol when the alternative (Kelly et al.[43]) value for the proton solvation energy is used instead.

By way of comparison, the SM12 solvation model[47] that was parametrized using the same experimental data (with the Kelly et al. proton reference) affords MUEs for ions that are considerably larger, ranging from 5.7 to 6.8 kcal/mol in DMSO and from 5.5 to 6.1 kcal/mol in CH_3CN, depending on the functional and basis set that is used to describe the solute. The MUEs for CMIRS v. 1.1 are thus considerably smaller, despite the use of fewer empirical parameters as compared to SM12. It does bear mentioning that SM12 uses more parameters but is designed to be "universal" (in the sense of being applicable to all solvents)[4] and thus might be more competitive if parametrized on a per solvent basis, as is CMIRS. Nevertheless, for the five solvents currently available in CMIRS v. 1.1, the accuracy of this model seems to compare well with the best available QM implicit solvation models.

Last, we come to water, for which we find that the A and B parameters are smaller in magnitude as compared to v. 1.0, but the C and D parameters are larger, except for the case when $\rho_0 = 0.0005$ au, while γ is about the same. Using $\rho_0 = 0.0010$ au, we are able to achieve a MUE of 0.78 kcal/mol for neutral solutes, 2.36 kcal/mol for ions, and 1.25 kcal/mol considering all solutes. The actual data are plotted against experimental values in Figure 6.

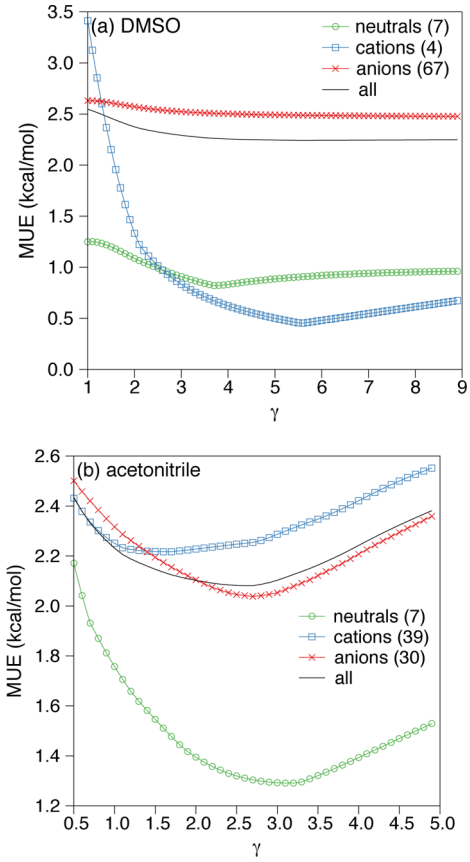


Figure 4. Mean unsigned errors for (a) DMSO and (b) CH_3CN, using $\rho_0 = 0.001$ au, B3LYP/6-31+G*, and Fawcett's value[44] for the proton solvation energy. Numbers in parentheses in the legend indicate the number of data points in each case.

Considering all of the data, we recommend the parameters optimized using $\rho_0 = 0{:}0010$ au and the Kelly et al. proton solvation energies, for benzene, cyclohexane, acetonitrile, and water. For DMSO, we recommend the parameters optimized using $\rho_0 = 0{:}0005$ au and the Kelly et al. proton reference. For DMSO, additional data for neutral and cationic solutes would likely increase the accuracy of the parametrization, or at least boost confidence in its robustness.

III.B. Robustness of the Parameterization. Following Pomogaeva and Chipman,[22−24] we have performed some additional analysis of the parametrization of the model. First, in order to ascertain the robustness of the parametrization, we took the data set for each solvent and selected half of the points (at random) to serve as a fitting set, with the other half reserved for evaluation, then reversed the roles of the two data sets and refit the parameters. This exercise (including refitting the parameters) was repeated 10 000 times for each solvent, with MUEs that are slightly different in each case because the subsets are chosen randomly. All parameters including γ were optimized simultaneously using the Nelder−Mead algorithm in SciPy[41] to minimize the MUE, so that scanning of γ profiles can be avoided in order to facilitate the very large number of parameter fits. Having already optimized the CMIRS parameters using the RMSE criterion, we now know that the MUE(γ) profiles are stable in the vicinity of the optimal value of γ (see Figure 3), so the MUE optimization criterion is not problematic. Results for the *change* in the MUE (with respect to the values reported in Table 2 that were obtained by fitting to

DOI: 10.1021/acs.jctc.6b00644
J. Chem. Theory Comput. 2016, 12, 4338−4346

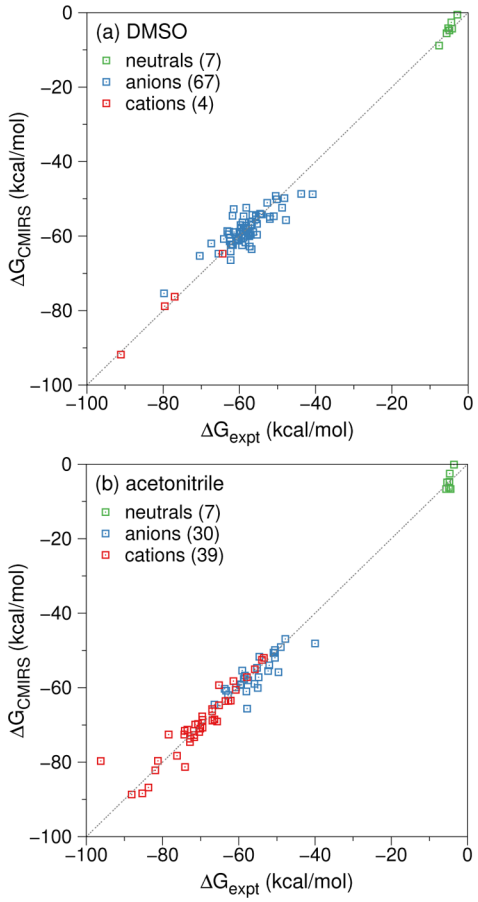


Figure 5. CMIRS v. 1.1 solvation energies versus experimental values for (a) DMSO and (b) CH_3CN, using ρ_0 = 0.001 au, B3LYP/6-31+G*, and Fawcett's value[44] of the proton solvation energies. Numbers in parentheses in the legend indicate the number of data points in each case.

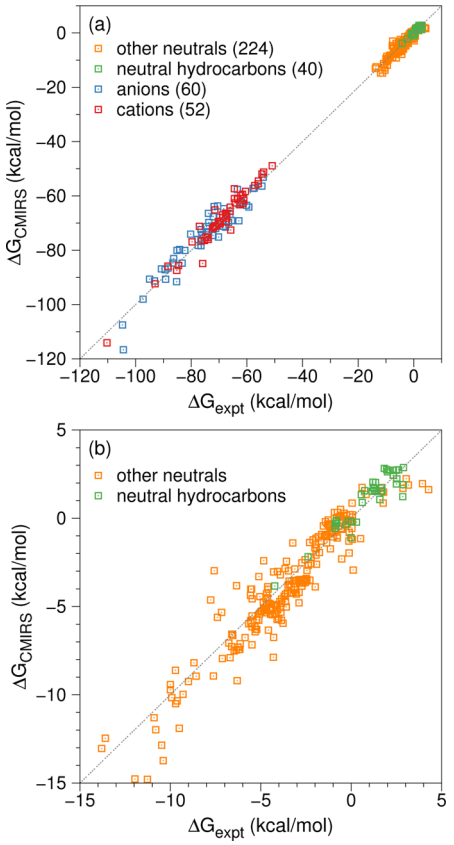


Figure 6. CMIRS v. 1.1 solvation energies in water, versus experimental values for (a) the entire data set of solutes and (b) neutral solutes only. An isocontour value ρ_0 = 0.0010 au was used along with B3LYP/6-31+G* for the solute, and the numbers in parentheses in the legend indicate the number of data points in each case.

the entire data set), along with the standard deviation in this change across all 10 000 refitted sets of parameters, are reported in Table 3.

For three of the five solvents, changes in the MUE are quite small across these 10 000 reparameterizations, which inspires confidence in the robustness of the parametrization. Relatively larger changes are observed for DMSO and CH_3CN, and we strongly suspect that this arises from the somewhat unusual nature of the data sets for these two solvents. Unlike water, for example, where the data set contains 112 ionic solutes and 264 neutral solutes, there are only seven neutral solutes in the data set for either DMSO or CH_3CN. For DMSO, there are 67 anions but only four cations, whereas for CH_3CN there are 30 anions and 39 cations. Especially for DMSO, many of the 10 000 instances of our randomized procedure afford fitting sets containing only anions, which lead to larger MUEs for the neutral solutes, and this seems to primarily affect the need to reoptimize the nonlinear γ parameter in the FESR model. The deviations reported in Table 3 reflect reoptimization of all parameters, but only for DMSO does reoptimization of γ matter significantly. If only the linear parameters are reoptimized for this solvent, we find that the MUE changes by a significantly larger value, 1.09 ± 8.94 kcal/mol.

For CMIRS v. 1.0, the quantities ΔG_{disp} and ΔG_{exch} are found to be highly correlated, e.g., with a correlation coefficient of R^2 = 0.99 for water.[23] Despite the change in the implementation of

Table 3. Change in MUE upon Refitting Parameters for 10 000 Randomly Selected Subsets Consisting of Half of the Available Experimental Data

solvent	ΔMUE (kcal/mol)
cyclohexane	0.08 ± 0.06
benzene	0.16 ± 0.11
DMSO	0.43 ± 0.81
CH_3CN	0.48 ± 0.33
water	0.13 ± 0.10

ΔG_{disp}, we obtain essentially the same correlation coefficient for CMIRS v. 1.1. This implies that significant cancellation of the attractive dispersion interaction by the repulsive exchange interaction remains a feature of the corrected model, as quantified above in the specific case of cyclohexane, for which both the dispersion and exchange interactions are relatively large. This is the case despite that fact that the integral in eq 3 that defines ΔG_{disp} becomes more attractive in the present implementation, by a sizable amount for larger solutes. Compensation is observed in the refitting of the A and B parameters.

Since our implementation of CMIRS is exactly the same as v. 1.0 except for this dispersion term, the high correlation between dispersion and exchange suggests that the parameters in ΔG_{FESR} should not change significantly in the new implementation, and indeed this is what we observe. For

DOI: 10.1021/acs.jctc.6b00644
J. Chem. Theory Comput. 2016, 12, 4338−4346

calculations based on Fawcett's proton reference, ΔG_{FESR} for the four cationic solutes in DMSO equals −5.59, −6.65, −5.77, and −7.84 kcal/mol in our implementation, versus −5.69, −6.64, −5.86, and −7.70 kcal/mol in v. 1.0. These changes are essentially negligible, while the dispersion and exchange energies change by 25%. Similar remarks hold for water and for CH_3CN, and Figure 7 demonstrates that the v. 1.0 and v. 1.1 values of ΔG_{FESR} are essentially identical, despite the reparameterization. This comparison speaks to the robustness of the FESR model, at least for the solutes considered here.

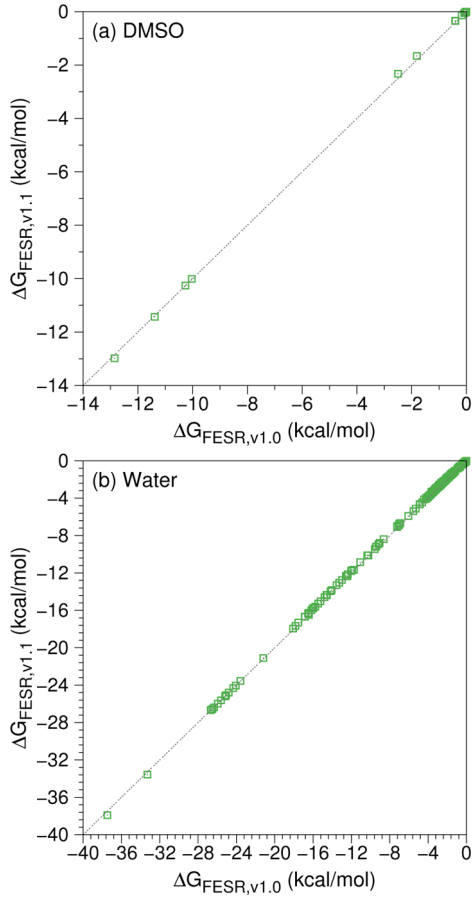


Figure 7. Comparison of the value of the ΔG_{FESR} term in CMRIS v. 1.0 versus v. 1.1, for (a) DMSO and (b) water. An isocontour value ρ_0 = 0.001 au was used along with B3LYP/6-31+G* for the solute.

IV. SUMMARY

We have corrected an error in the implementation of the CMIRS solvation model, which necessitated refitting the five empirical parameters that define this model, in each of the five solvents (water, DMSO, CH_3CN, benzene, and cyclohexane) for which it has been parametrized. The correction alters the dispersion energy by several kilocalories per mole in some cases, but upon refitting the parameters we find that the overall error statistics are only slightly improved, relative to CMIRS v. 1.0. The marginal improvement arises from a compensating effect in the exchange interaction when the parameters of are refit. Excellent correlation between the magnitude of the dispersion and exchange energies, observed originally for CMIRS v. 1.0, is preserved here, and the value of the single nonlinear parameter in the model changes little upon reoptimization, suggesting that the model captures much of the essential physics with relatively few parameters. For solutes described at the B3LYP/6-31+G* level, we achieve MUEs of <0.7 kcal/mol for benzene and cyclohexane, < 1.5 kcal/mol for water, and <2.8 kcal/mol for DMSO and CH_3CN.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00644.

Energy components for each molecule in the database using the optimized CMIRS v. 1.1 parameters (XLSX)

Energy components for each molecule in the database using the optimized CMIRS v. 1.1 parameters (TXT)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: herbert@chemistry.ohio-state.edu.

Notes

The authors declare the following competing financial interest(s): J.M.H. serves on the Board of Directors of Q-Chem, Inc.

■ ACKNOWLEDGMENTS

We thank Daniel Chipman and Anna Pomagaeva for assistance in confirming the technical issue documented here, and for suggesting the use of RMSE rather than MUE as an error metric. This work was supported by National Science Foundation grant no. CHE-1300603 and by a Camille-Dreyfus Teacher-Scholar Award to J.M.H., who also acknowledges a fellowship from the Alexander von Humboldt Foundation. Calculations were performed at the Ohio Supercomputer Center.[48]

■ 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) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3093.
(4) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760−768.
(5) Klamt, A. *WIREs Comput. Mol. Sci.* **2011**, *1*, 699−709.
(6) Tomasi, J.; Mennucci, B.; Cancès, E. *J. Mol. Struct.: THEOCHEM* **1999**, *464*, 211−226.
(7) Chipman, D. M. *J. Chem. Phys.* **1999**, *110*, 8012−8018.
(8) Chipman, D. M. *J. Chem. Phys.* **2000**, *112*, 5558−5565.
(9) Herbert, J. M.; Lange, A. W. In *Many-Body Effects and Electrostatics in Biomolecules*; Cui, Q., Ren, P., Meuwly, M., Eds.; Pan Stanford, 2016; Chapter 11, pages 363−416.
(10) Chipman, D. M. *Theor. Chem. Acc.* **2002**, *107*, 80−89.
(11) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J. *Acc. Chem. Res.* **2009**, *42*, 489−492.
(12) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 4538−4543.
(13) Floris, F. M.; Tomasi, J.; Ahuir, J. L. P. *J. Comput. Chem.* **1991**, *12*, 784−791.
(14) Amovilli, C.; Mennucci, B. *J. Phys. Chem. B* **1997**, *101*, 1051−1057.
(15) Amovilli, C.; Barone, V.; Cammi, R.; Cancès, E.; Cossi, M.; Mennucci, B.; Pomelli, C. S.; Tomasi, J. *Adv. Quantum Chem.* **1999**, *32*, 227−261.
(16) Curutchet, C.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2001**, *22*, 1180−1193.
(17) Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Orozco, M.; Luque, F. J. *J. Mol. Struct.: THEOCHEM* **2005**, *727*, 29−40.

DOI: 10.1021/acs.jctc.6b00644
J. Chem. Theory Comput. 2016, 12, 4338−4346

(18) Curutchet, C.; Orozco, M.; Luque, F. J.; Mennucci, B.; Tomasi, J. *J. Comput. Chem.* **2006**, *27*, 1769−1780.
(19) Weijo, V.; Mennucci, B.; Frediani, L. *J. Chem. Theory Comput.* **2010**, *6*, 3358−3364.
(20) Pomogaeva, A.; Chipman, D. M. *J. Chem. Theory Comput.* **2011**, *7*, 3952−3960.
(21) Pomogaeva, A.; Thompson, D. W.; Chipman, D. M. *Chem. Phys. Lett.* **2011**, *511*, 161−165.
(22) Pomogaeva, A.; Chipman, D. M. *J. Phys. Chem. A* **2013**, *117*, 5812−5820.
(23) Pomogaeva, A.; Chipman, D. M. *J. Chem. Theory Comput.* **2014**, *10*, 211−219.
(24) Pomogaeva, A.; Chipman, D. M. *J. Phys. Chem. A* **2015**, *119*, 5173−5180.
(25) Cancès, E.; Mennucci, B. *J. Chem. Phys.* **2001**, *114*, 4744−4745.
(26) Lange, A. W.; Herbert, J. M. *Chem. Phys. Lett.* **2011**, *509*, 77−87.
(27) Vydrov, O. A.; Van Voorhis, T. *Phys. Rev. Lett.* **2009**, *103*, 063004:1−4.
(28) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2011−2033.
(29) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 3649−3659.
(30) 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.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(31) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kús, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L., III; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A., Jr.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Pieniazek, P. A.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sergueev, N.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Vanovschi, V.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A., III; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F., III; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Mol. Phys.* **2015**, *113*, 184−215.
(32) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098−16104.
(33) Zhan, C.-G.; Bentley, J.; Chipman, D. M. *J. Chem. Phys.* **1998**, *108*, 177−192.
(34) Chipman, D. M.; Dupuis, M. *Theor. Chem. Acc.* **2002**, *107*, 90−102.
(35) Gor'kov, L. P.; Pitaevskii, L. P. *Sov. Phys. Dokl* **1964**, *8*, 788−790.
(36) Herring, C.; Flicker, M. *Phys. Rev.* **1964**, *134*, A362−A366.
(37) McWeeny, R. *Methods of Molecular Quantum Mechanics*, 2nd ed.; Academic Press: New York, 1992.
(38) Corcelli, S. A.; Skinner, J. L. *J. Phys. Chem. A* **2005**, *109*, 6154−6165.
(39) Smith, J. D.; Cappa, C. D.; Wilson, K. R.; Cohen, R. C.; Geissler, P. L.; Saykally, R. J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 14171−14174.
(40) 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.
(41) Jones, E.; Oliphant, T.; Peterson, P.; et al. *SciPy: Open source scientific tools for Python*, 2001.
(42) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(43) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 408−422.
(44) Fawcett, W. R. *Langmuir* **2008**, *24*, 9868−9875.
(45) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. J., Jr. *J. Phys. Chem. A* **1998**, *102*, 7787−7794.
(46) Donald, W. A.; Williams, E. R. In *Electroanalytical Chemistry: A Series of Advances*; Bard, A. J., Zoski, C., Eds.; CRC Press: Boca Raton, FL, 2014; Vol. *25*, chapter 1, pp 1−32.
(47) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 609−620.
(48) Ohio Supercomputer Center. http://osc.edu/ark:/19495/f5s1ph73 (accessed 24 July, 2016).

DOI: 10.1021/acs.jctc.6b00644
J. Chem. Theory Comput. 2016, 12, 4338−4346