Performance of SM6, SM8, and SMD on the SAMPL1 test set for the prediction of small-molecule solvation free energies

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · MAY 2009

Impact Factor: 3.3 · DOI: 10.1021/jp809094y · Source: PubMed

3 AUTHORS, INCLUDING:

Christopher J Cramer
University of Minnesota Twin Cities
532 PUBLICATIONS **23,502** CITATIONS
SEE PROFILE

Donald Truhlar
University of Minnesota Twin Cities
1,342 PUBLICATIONS **81,884** CITATIONS
SEE PROFILE

Performance of SM6, SM8, and SMD on the SAMPL1 Test Set for the Prediction of Small-Molecule Solvation Free Energies†

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

Department of Chemistry and Supercomputing Institute, 207 Pleasant St. SE, University of Minnesota, Minneapolis, Minnesota

Received: October 14, 2008; Revised Manuscript Received: November 24, 2008

The SM6, SM8, and SMD quantum mechanical aqueous continuum solvation models are applied to predict free energies of aqueous solvation for 61 molecules in the SAMPL1 test set described elsewhere (Guthrie. *J. Phys. Chem. B* **2009**, *113*, 4501−4507). For direct comparison to other models, frozen geometries, provided by Guthrie, were used together with the M06-2X density functional and the 6-31G(d) basis set. For the bulk electrostatic component of the solvation free energy, SM6 and SM8 employ a generalized Born model that uses polarized discrete partial atomic charges to model the electron density, with these charges being calculated by the CM4 and CM4M class IV charge models, respectively; SMD uses the polarized continuous quantum mechanical charge density. If five sulfonylureas are removed from the SAMPL1 set, the root-mean-square deviations (RMSDs) of SM6, SM8, and SMD on the remaining 56 molecules are 2.4, 2.6, and 2.5 kcal mol^{-1}, respectively. The SM6, SM8, and SMD RMSDs on the five sulfonylureas are 14.2, 12.6, and 11.1 kcal mol^{-1}, respectively; however, we suggest that the uncertainty in the target solvation free energies for these molecules may be quite large.

Introduction

The accurate prediction of solvation free energies continues to be a challenging task for explicit[1−3] and implicit[3−5] theoretical solvation models.[6] Ongoing efforts are motivated in part because of the considerable importance of solvation as it influences such phenomena as changes in reaction rates and mechanisms,[6−8] the hydrophobic effect,[9] chromatographic retention behavior,[10] interfacial concentration and transport,[11,12] protein folding,[13] protein−ligand binding,[14] and pharmaceutical (or toxin) bioavailability.[15−19] With these last two topics particularly in mind, in early 2008 Nicholls et al.[20] offered as a standard test set 17 small molecules for which the free energies of aqueous solvation had been determined; a noteworthy feature of the set was the presence of drug-like functionality in many of the molecules. Considering various explicit and implicit solvent models combined with a molecular mechanics description for the solutes, Nicholls et al. reported root-mean-square deviations (RMSDs) ranging from 1.3 to 2.6 kcal mol^{-1}.[20] We applied the SM8 generalized Born solvation model,[21] in its default implementation using the M06-2X[22] density functional, the 6-31G(d)[23] basis set, and the Charge Model 4/M06-suite (CM4M)[24] charge model with gas-phase geometries optimized with the M06-2X[22] density functional, and reported an rmsd of 1.1 kcal mol^{-1} for the same data set.[25]

The testing of different models against common test sets is a particularly valuable exercise that facilitates comparison of their relative strengths and weaknesses. Guthrie,[26] as part of the present coordinated series of articles, compiled a new test set of 63 molecules, many of which are agricultural pesticides, for which he determined target free energies of aqueous solvation based on a careful search of relevant experimental literature data (see below for additional discussion). The geometries of these molecules were provided as Cartesian coordinates to a number of groups active in solvation model development with the request that they compute solvation free energies by their various protocols for comparison in a special collection of articles in this journal. The Associate Editor in charge invited the current authors to participate as well, and we here report our results as obtained from some of our most recently developed SM*x* quantum mechanical continuum solvation models.

† Part of the special section "Calculation of Aqueous Solvation Energies of Drug-Like Molecules: A Blind Challenge".

Computational Methods

In a continuum solvation model, the free energy of interaction between a solute and a solvent may be considered to be composed of bulk electrostatic contributions and other components, although the partition into the two components is not unique. The latter components include such physical phenomena as cavitation, dispersion, solvent structure in the first solvation shell, the conversion of solute rotations into librations, the partial covalent component of any solute−solvent hydrogen bonding, and the breakdown of bulk electrostatics in the first solvation shell, and they also correct in part for the nonuniqueness of the partitioning into bulk electrostatics and other effects. All of the models employed in this paper adopt the SM*x* formalism[27−29] in which an atomic surface tension is assigned semiempirically to each atom in a molecule based on its atomic number and its intramolecular environment (i.e., based on its atomic number and the solute geometry). The nonbulk-electrostatic contribution G_{CDS} to the full free energy of solvation is then computed as

$$G_{CDS} = \sum_k A_k \sigma_k \quad (1)$$

where CDS denotes cavity-dispersion-solvent-structure, A_k is the solvent-accessible surface area of atom k, and σ_k is its geometry-dependent atomic surface tension. The geometry dependence is controlled by a series of geometrical functional

10.1021/jp809094y CCC: $40.75
Published on Web 03/02/2009

forms, and the parameters involved in the atomic surface tensions are optimized to reproduce experimental free energies of solvation for nonelectrolytes; as a consequence, they also correct semiempirically for systematic errors that are correctable with the functional forms chosen. Our most recent training set, used for SMD,[30] includes data for 274 neutral solutes and 112 ions in water, 2346 neutral solute data and 220 ionic data for nonaqueous solvents, and 143 transfer free energies from water to organic solvents (although the present paper is concerned only with aqueous solvents, the presence of transfer free energies means that the final parameters in aqueous solution have some dependence on all the data in all the solvents). The SM8 model[21] was parametrized to the same data plus 21 data for ionic clusters in nonaqueous solution. The SM6 model[31] was parametrized to data for 273 neutral solutes and 112 ions in water (tetramethylsilane was not used for the SM6 parametrization).

For the electrostatic component of solvation, which depends upon the self-consistent reaction field induced in the solvent by the charge distribution of the solute, we consider here two alternative formalisms. Most SM*x* models[27−29] compute the bulk-electrostatic polarization component of the free energy of solvation G_P from the generalized Born (GB)[27,32−37] approximation according to

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

where ε is the solvent dielectric constant, the summation runs over all atoms k and k' having partial atomic charges q_k and $q_{k'}$, and $\gamma_{kk'}$ is an effective Coulomb integral that has units of inverse distance and that depends on geometry-dependent atomic radii. The geometry-dependent atomic radii depend on intrinsic atomic radii, which are parameters, and on a dielectric descreening algorithm[36,38] based on the Coulomb-field approximation. The details of the SM8 solvation model,[21] and its key underlying GB algorithms,[38−40] are available in the literature and will not be recapitulated here. For aqueous solution, the SM8 model is very similar in construction to the earlier generation SM6[31] model. The key difference between the two, however, is that SM6 is defined only for aqueous solvation while SM8 is defined for general solvents. Because water is so important, there may be advantages to using a less general (more specifically parametrized) solvation model for this solvent, so we consider SM6 in the present paper in addition to SM8.

We have also developed an SM*x* model that we call SMD.[30] The "D" in the name stands for "density" and indicates that we do not use partial charges and the self-consistent GB formalism to compute the bulk-electrostatic component of the solvation free energy, but we instead solve the nonlinear Poisson equation (NPE) self-consistently for the continuous quantum mechanical charge density. The cavity for the nonlinear Poisson approximation is constructed in much the same way as that for the GB approximation. We adopt the IEF-PCM protocol[41] with an all-atom molecular cavity constructed from a set of atomic radii optimized to reproduce ionic solvation free energies in water, methanol, acetonitrile, and dimethyl sulfoxide[42,43] (the much greater sensitivity of ionic solvation free energies to atomic radii motivates the optimization of the cavity radii against such data). We continue to employ our usual surface-tension-based approach to compute the non-bulk-electrostatic component of the free energy of solvation. Because the separation of the full free energy of solvation into bulk-electrostatic and non-bulk-electrostatic components is not well defined[44] and because GB and NPE predictions for the bulk-electrostatic component can differ markedly for molecules decorated with polar functionality (although this dependence is reduced by using different atomic radii in the two formalisms), we consider it valuable to compare predictions from these alternative electrostatic formalisms on this demanding test set.

When restricted to water as solvent, we note that the SMD model and the MST solvation model of Soteras et al.[45] are functionally very similar to one another. The MST model combines IEF-PCM electrostatics with a cavitation term based on scaled-particle theory[46−48] (not separately present in SMD but included implicitly in eq 1) and a van der Waals term computed from water-specific atomic or group surface tensions precisely as in eq 1. We did not test the aqueous MST model on the present SAMPL1 set; prior studies on small-molecule training sets have generally indicated good agreement between MST and SM*x* models for aqueous free energies of solvation,[45,49−52] but there is functionality in the SAMPL1 set that is poorly represented in most prior training sets appearing in the literature, so it would be interesting to see results from the MST model in the future.

For completeness, we note that IEF-PCM or PCM electrostatics (PCM solves the same NPE problem as IEF-PCM but does not use an integral-equation formalism[5]) have been combined with various other approaches to compute non-bulk-electrostatic contributions to predict aqueous free energies of solvation,[53−55] but applications of these models to test sets other than their original training sets, which were composed exclusively of small neutral solutes, have not been described. One exception is a particular protocol for the construction of PCM electrostatic cavities (united atom Hartree−Fock (UAHF)) described by Barone et al.[56] and designed to provide accurate aqueous solvation free energies when combined with cavitation free energies from scaled-particle theory and dispersion-repulsion free energies from classical pairwise potentials. Subsequent evaluation of the PCM(UAHF) approach on a larger test set than that used for parametrization indicated substantial loss of accuracy associated with low diversity in the original training set.[21]

All the free energies of solvation discussed here are standard-state free energy changes that correspond to a 1 M ideal gas and a 1 M ideal infinitely dilute solution. This convention eliminates any change in the translational (liberational) entropy upon solvation as discussed by Ben-Naim.[57] All molecular geometries were taken as received from the work of Guthrie[26] and are available in his Supporting Information.

Software

The SM8 solvation model is included in the GAMESS-PLUS,[58] MN-GSM,[59] and Q-Chem[60] software packages, the SMD model is in the GESOL[61] program, and the SM6 solvation model is included in the GAMESSPLUS, MN-GSM, HONDOPLUS,[62] SM*x*GAUSS,[63] Jaguar,[64] and Q-Chem software packages. The present calculations were all carried out with MN-GSM.

Results and Discussion

Guthrie[26] describes an original test set of 63 molecules (see Figure 1 of his paper for index names, common names, and structures; we will employ the same index and common names here) and reports results from selected other research groups for a subset of 56 of these molecules, referring to "various problems in data preparation" as motivating the removal of seven of the data. Of these seven, in one case, cup08062 (4-

dimethylamino-4′-nitroazobenzene), the supplied geometry file did not include the nitro group, so we removed this molecule from the original set as well. In another case, cup08042 (oxamyl), one[65] of the two sources cited for experimental data failed to contain any mention of the subject compound (and we did not have ready access to the other[66]), so we removed this molecule as well. In the remaining five cases (cup08037, methomyl; cup08058, 4-amino-4′-nitroazobenzene; cup08059, 1-amino-4-anilinoanthraquinone; cup08060, 1,4,5,8-tetraminoanthraquinone; and cup08061, 1-aminoanthraquinone), we did not identify any obvious reason to discard the molecules from consideration, and our predictions were in generally good agreement with the reported experimental values. Thus, we maintain these five molecules in our discussion, but treat them as a separate class as there may be some reason for caution in interpreting these predictions. Finally, five of the molecules in the SAMPL1 test set are sulfonylureas (cup08012, bensulfuron; cup08020, chlorimuron ethyl; cup08039, metsulfuron methyl; cup08051, sulfometuron methyl; cup08054, thifensulfuron). For reasons discussed in more detail below, we consider these molecules as a separate subset as well. Thus, we have three subsets in our discussion: the 5 sulfonylureas, the 5 of 7 molecules not used by original invitees (the other two we discard as well), and the 51 molecules remaining after removing these special cases.

Predictions for the three subsets defined above are provided in Table 1 together with the experimental values provided by Guthrie.[26] The mean signed deviation (MSD), mean unsigned deviation (MUD), and RMSD values for the various models over the different subsets are also provided. We do not list the error estimates of Guthrie for the experimental values. In most instances, the solvation free energy targets were derived from separately reported solubilities and vapor pressures (expressed in appropriate standard-state free energy units[67]). In the many cases where the original experimental reports failed to provide error estimates, Guthrie allowed for errors of up to 1 order of magnitude in measured equilibrium constants, leading to suggested error bars of ± 1.4 or ± 1.9 kcal mol^{-1}, depending on whether this error approximation applied to only one of the two individual measurements or to both.[26]

Over the 51-molecule subset, the rmsds of SM6, SM8, and SMD are 2.4, 2.7, and 2.6 kcal mol^{-1}, respectively. The RMSD as an error indicator tends to amplify the impact of a small number of large outliers, so it is useful to consider other error measures as well. Thus, for instance, the MUDs of SM6, SM8, and SMD over the 51-molecule subset are 1.6, 2.1, and 2.0 kcal mol^{-1}, respectively. The significant differences between the MUDs and RMSDs can be attributed to a small number of molecules showing large errors. For the SM6 model, cup08029 (endosulfan alpha) and cup08040 (nitralin) show significant errors. In addition to these same two molecules, cup08023 (dialifor) and cup08031 (ethion) also are problematic for the SM8 model. For SMD, it is cup08001 (nitroglycol), cup08023, cup08029, cup08031, and cup08049 (pyrazon) that prove to be the large outliers. As a final error measure, we note that SM6, SM8, and SMD are within 1 kcal mol^{-1} of the experimental estimate for 24, 14, and 17 of the 51 molecules, respectively, and within 2 kcal mol^{-1} of the experimental estimate for an additional 14, 18, and 13 molecules, respectively.

With the exception of cup08060, the three SM*x* models make predictions of roughly similar accuracy within the five-molecule set distinguished as questionable by Guthrie. Indeed, in 7 of 15 instances the predictions in this subset are within 1 kcal mol^{-1} of the experimental value. If such agreement is purely fortuitous, because there are errors in the geometries or the experimental data, then such small errors in the predictions are a remarkable coincidence.

Finally, we consider the sulfonylureas. In every instance, all three SM*x* models predict solvation free energies in fairly good agreement with one another, but much more negative than the derived experimental solvation free energies. Guthrie[26] notes that, of these five sulfonylureas, cup08012, cup08020, and cup08054 were also significant outliers for all of the groups originally invited to participate in the SAMPL1 challenge. Examination of the original literature suggests that the experimentally derived solvation free energies for the sulfonylureas may be insufficiently negative because of flaws in the methodology used to determine their vapor pressures. As the compounds are highly nonvolatile, their vapor pressures were measured by assessing slow mass loss from solid samples heated to rather high temperatures (Knudsen effusion, performed over temperature ranges from as low as 110 to as high as 170 °C depending on individual compound) and then extrapolating back to room temperature.[68] In addition to uncertainties associated with the extrapolation procedure (which are large), there is the potential for sample decomposition at high temperature. When such decomposition creates smaller molecular fragments, with correspondingly higher vapor pressures, the mass loss, and hence the vapor pressure, will be measured to be unrealistically high. Moreover, the authors of the original study note very large differences in measured values from different pesticide samples.[68] On that basis, we consider the experimental solvation free energy targets assigned for these compounds to be suspect.

To be balanced, it is certainly possible that the SM*x* models overestimate the portion of the solvation free energy associated with the sulfonylurea fragment. Thus, we note that a large outlier in the 51-molecule subset is cup08029, which is a sulfite diester that is also predicted by the SM*x* models to be substantially too well solvated. On the other hand, this molecule, too, is listed by Guthrie[26] as one for which all of the originally invited models performed poorly. There is no question that the training sets for most solvation models have very poor representations of oxidized sulfur functionalities—there are none for SM6, SM8, and SMD—so it is possible that new parametrizations including, for instance, cup08029 (the data for which[26] seem reasonably robust) would be able to address this issue with little or no loss in accuracy for other functionalities. Additional examples would, of course, also be desirable, but the very low vapor pressures of the sulfonylureas suggests that they may not be ideal choices.

A few other trends present themselves in the SM*x* data. For example, the SMD model appears to underestimate the solvation free energy associated with nitro groups (which dominate the functionality in, for example, cup08001−08006 and cup08021), although in this case the SMD aqueous training set does include seven nitroaliphatics and nitroaromatics. Additionally, the SM6 model appears to be more accurate for compounds containing thionothiophosphate ester functionality (e.g., cup08023, cup08031, and cup08036). This may reflect the parametrization of SM6 only for aqueous data, while SM8 was designed to be a universal solvation model, i.e., its training set included many solvation free energies in organic solvents, too. Thus, the SM6 and SM8 aqueous trainings sets both included five phosphate esters, seven thiophosphate esters, and one thiophosphonate ester, but the SM8 aqueous parameters were optimized over a training set that made use of an additional eight transfer free energies for isopropyl methylphosphonofluoridate between water and various organic solvents.

TABLE 1: Aqueous Solvation Free Energies (kcal mol^{-1}) from Experiment and from the SM6, SM8, and SMD Continuum Solvation Models Using Geometries Provided by Guthrie

	formula[a]												
molecule[a]	*m*	*n*	*o*	*p*	*q*	*r*	*s*	*t*	*u*	SM6	SM8	SMD	expt[b]
51-molecule subset													
cup08001	4	2	2	6	0	0	0	0	0	−2.9	−2.2	−0.8	−5.7
cup08002	6	3	2	6	0	0	0	0	0	−3.2	−2.6	−1.3	−5.0
cup08003	9	4	1	3	0	0	0	0	0	−0.7	−0.6	0.3	−2.1
cup08004	9	4	1	3	0	0	0	0	0	−0.4	−0.3	0.7	−1.8
cup08005	9	4	1	3	0	0	0	0	0	−0.3	−0.2	0.6	−1.9
cup08006	5	2	1	4	0	0	0	0	0	−6.4	−6.0	−5.4	−8.2
cup08007	20	14	1	2	0	0	0	1	0	−6.1	−6.3	−8.4	−8.2
cup08008	14	7	2	2	0	0	1	0	0	−8.5	−8.5	−6.8	−9.8
cup08009	17	9	5	0	0	0	1	0	0	−9.5	−10.5	−9.6	−7.7
cup08010	12	10	3	3	0	1	2	0	0	−9.9	−12.2	−13.8	−10.0
cup08011	16	13	3	4	3	0	0	0	0	−1.4	−1.7	−0.6	−3.5
cup08013	13	9	2	2	0	0	0	0	1	−9.6	−10.3	−10.8	−9.7
cup08014	8	9	1	2	0	0	1	3	0	−8.2	−7.9	−9.0	−9.0
cup08015	11	12	1	2	0	0	0	0	0	−11.5	−11.7	−10.1	−9.5
cup08016	15	12	1	3	0	0	0	0	0	−12.2	−12.3	−10.9	−9.6
cup08017	16	11	0	2	0	1	3	1	0	−5.4	−7.8	−7.8	−6.5
cup08018	6	10	0	0	0	0	0	8	0	−3.0	−2.7	−4.4	−3.4
cup08019	14	12	0	4	0	1	0	3	0	−7.3	−5.2	−6.2	−7.1
cup08021	0	1	1	2	0	0	0	3	0	−1.6	−1.3	1.5	−1.5
cup08022	11	9	1	3	0	1	1	3	0	−4.8	−6.0	−4.7	−5.0
cup08023	17	14	1	4	0	1	2	1	0	−8.8	−11.1	−12.2	−5.7
cup08024	21	12	2	3	0	1	1	0	0	−7.3	−9.1	−7.8	−6.5
cup08025	6	8	0	3	0	0	0	2	0	−8.0	−7.9	−6.8	−9.9
cup08026	3	7	1	0	0	0	0	2	0	−4.0	−3.3	−2.8	−4.7
cup08027	13	11	4	4	3	0	0	0	0	−3.7	−3.7	−2.9	−5.7
cup08028	12	10	2	5	0	0	0	0	0	−9.7	−9.6	−8.3	−6.2
cup08029	6	9	0	3	0	0	1	6	0	−15.3	−13.4	−10.5	−4.2
cup08030	8	12	0	1	0	0	0	6	0	−6.3	−5.6	−4.7	−5.5[c]
cup08031	22	9	0	4	0	2	4	0	0	−5.8	−11.8	−12.0	−6.1
cup08032	12	9	2	1	0	0	0	0	0	−9.5	−10.3	−9.4	−9.1
cup08033	5	10	0	0	0	0	0	7	0	−2.1	−1.8	−2.3	−2.6
cup08034	14	9	0	1	0	0	0	0	0	−4.2	−4.5	−4.6	−5.2
cup08035	6	6	0	0	0	0	0	6	0	−3.8	−3.6	−7.2	−5.4
cup08036	19	10	0	6	0	1	2	0	0	−9.1	−11.2	−9.3	−8.2
cup08038	10	8	1	5	0	1	1	0	0	−7.8	−8.6	−6.1	−7.2
cup08040	19	13	3	6	0	0	1	0	0	−14.2	−13.1	−10.9	−8.0
cup08041	5	3	1	4	0	0	0	0	0	−5.4	−5.1	−3.5	−6.0
cup08043	14	10	1	5	0	1	1	0	0	−6.5	−7.9	−6.3	−6.7
cup08044	21	10	1	1	0	0	1	0	0	−2.7	−3.0	−3.1	−3.6
cup08045	17	7	0	2	0	1	3	0	0	−4.1	−6.8	−7.2	−4.4
cup08046	16	14	3	4	3	0	0	0	0	−2.0	−2.1	−1.4	−2.5
cup08047	19	10	5	0	0	0	1	0	0	−7.1	−8.3	−7.9	−8.4
cup08048	9	9	1	1	0	0	0	2	0	−8.5	−8.6	−7.6	−7.8
cup08049	8	10	3	1	0	0	0	1	0	−12.8	−12.6	−11.1	−16.4
cup08050	12	7	5	0	0	0	0	1	0	−10.0	−11.1	−11.2	−10.2
cup08052	13	9	2	2	0	0	0	1	0	−8.9	−9.6	−9.2	−11.1
cup08053	19	10	5	0	0	0	1	0	0	−8.4	−9.4	−8.1	−6.7
cup08055	8	4	0	4	0	1	0	3	0	−10.1	−8.2	−11.6	−12.7
cup08056	16	13	3	4	3	0	0	0	0	−1.3	−1.5	−0.6	−3.3
cup08057	21	10	1	1	0	0	1	0	0	−3.2	−3.5	−3.6	−4.1
cup08063	18	11	4	2	0	0	0	0	0	−12.6	−13.8	−10.2	−9.4
MSE										**0.1**	**−0.3**	**0.2**	
MUE										**1.6**	**2.1**	**2.0**	
RMSD										**2.4**	**2.7**	**2.6**	
5-molecule subset													
cup08037	10	5	2	2	0	0	1	0	0	−11.5	−11.4	−9.5	−10.7
cup08058	10	12	4	2	0	0	0	0	0	−13.8	−13.1	−11.4	−11.2
cup08059	14	20	2	2	0	0	0	0	0	−8.7	−8.8	−8.4	−7.4
cup08060	12	14	4	2	0	0	0	0	0	−13.0	−12.6	−12.2	−8.9
cup08061	9	14	1	2	0	0	0	0	0	−8.9	−8.8	−7.8	−8.0
MSE										**−1.9**	**−1.7**	**−0.6**	
MUE										**1.9**	**1.7**	**1.2**	
RMSD										**2.3**	**2.0**	**1.6**	
sulfonylurea subset													
cup08012	18	16	4	7	0	0	1	0	0	−34.7	−29.6	−32.7	−17.2
cup08020	15	15	4	6	0	0	1	1	0	−27.0	−26.3	−23.1	−14.0
cup08039	15	14	5	6	0	0	1	0	0	−30.3	−30.0	−26.7	−15.5
cup08051	16	15	4	5	0	0	1	0	0	−30.0	−29.5	−25.9	−20.3
cup08054	13	12	5	6	0	0	2	0	0	−30.9	−30.1	−27.7	−16.2
MSE										**−14.0**	**−12.5**	**−10.6**	
MUE										**14.0**	**12.5**	**10.6**	
RMSD										**14.2**	**12.6**	**11.1**	

[a] The molecular formula is $H_mC_nN_oO_pF_qP_rS_sCl_t$. [b] See the work of Guthrie[26] for molecular structures, determination of experimental reference values, and experimental error estimates. [c] Guthrie[26] reports −4.8, but this appears to be a typo as −5.5 is the average of the three experimental values listed, all of which are more negative than −4.8.

The reasonably good agreement between predicted solvation free energies from SM6 and SMD merits additional discussion. Over all 61 molecules in Table 1, the average deviation between SM6 and SMD is 1.6 kcal mol^{-1}. However, the division between the non-bulk-electrostatic and bulk-electrostatic components can be quite different for the two different models. As a rule, SM6 predicts smaller bulk-electrostatic contributions, and SMD predicts larger ones.[44,52] As an extreme example, in sulfonylurea cup08012, the SM6 model predicts a bulk-electrostatic free energy of solvation of −26.0 kcal mol^{-1} while the SMD model predicts −41.5 kcal mol^{-1}. The average predictions of SM6 and SMD for this component of the solvation free energy over the full data set are −9.1 and −15.1 kcal mol^{-1}, respectively, with the SMD prediction always being more negative than the SM6 prediction. Nevertheless, after accounting for non-bulk-electrostatic effects, the two models provide predictions for the *full* free energy of solvation within 2 kcal mol^{-1} of one another. This illustrates again the degree to which it is important to ensure that models for computing the non-bulk-electrostatic contribution to the free energy of solvation are consistent with the electrostatic models and the degree to which it is not necessarily instructive to compare results from different solvation models if the analysis is restricted solely to either the bulk-electrostatic or non-bulk-electrostatic components.[44,52]

With respect to possible remaining sources of error, we note that many of the molecules in the SAMPL1 test set are floppy and may well exist as populations of multiple conformers at room temperature. In such instances, a proper averaging over gas-phase and solvated conformations should be done in order to compute the free energy of solvation.[6] However, from a practical standpoint, it is certainly true that the simplest way in which to estimate the solvation free energy is to employ a single, reasonable conformer. Another issue is the degree to which geometry relaxation would be expected to improve the solvation free energies of test set molecules. Such relaxation will always lead to the solvation free energy becoming more negative. As the mean signed errors for the various SM*x* models with frozen geometries are already fairly near zero, it seems unlikely that accounting for relaxation in solution would improve their accuracy (indeed, this result is expected since the parametrization of the models involved the use of gas-phase structures). Finally, the source of the molecular geometries was not detailed to us. It may be that the solvation free energies will be sensitive to geometry. Thus, for example, in our earlier study for the test set of Nicholls et al.[20] we found that our predictions improved slightly when M06-2X/6-31G(d,p) geometries were used in place of B3LYP/6-31G(d,p) geometries.[25]

We chose to explore the geometry issue in somewhat more depth for three cases, cup08032 (fenuron), cup08036 (malathion), and cup08040 (nitralin). These molecules were chosen because each has a reasonably large solvation free energy and together they present a number of the more polar functionalities in the test set. The results are summarized in Table 2. Reoptimization of the geometries of Guthrie[26] in the gas phase at the M06-2X/6-31G(d) level caused the gas-phase energies of cup08032, cup08036, and cup08040 to decrease by 2.2, 11.5, and 10.8 kcal mol^{-1}, respectively. While the latter two energy changes are not small in magnitude, the geometric changes involved were mostly minor adjustments in bond lengths—visualization revealed no distinguishable qualitative changes. In every case but one, the SM*x* solvation free energies for the M06-2X/6-31G(d) gas-phase geometries are predicted to be smaller in magnitude than those computed for the Guthrie geometries. For cup08032 and cup08040, however, the differences are very small, averag-

TABLE 2: SM*x* Aqueous Solvation Free Energies (kcal mol^{-1}) as a Function of Geometry

Molecule	geometry	SM6	SM8	SMD
cup08032	Guthrie	−9.5	−10.3	−9.4
	M06-2X/6-31G(d) (gas)	−9.3	−10.1	−9.0
	SM8/M06-2X/6-31G(d) (aq)		−10.9	
cup08036	Guthrie	−9.1	−11.2	−9.3
	M06-2X/6-31G(d) (gas)	−7.4	−9.5	−7.7
	SM8/M06-2X/6-31G(d) (aq)		−10.2	
cup08040	Guthrie	−14.2	−13.1	−10.9
	M06-2X/6-31G(d) (gas)	−13.9	−12.8	−11.0
	SM8/M06-2X/6-31G(d) (aq)		−13.4	

ing 0.2 kcal mol^{-1} (including the one case, SMD for cup08040, where the solvation free energy becomes *larger* in magnitude by 0.1 kcal mol^{-1}). For cup08036, on average, the solvation free energies become less negative by a larger 1.7 kcal mol^{-1}, and this reflects the degree to which the geometry and charge distribution, and hence the solvation free energy, of the thiono-thiophosphate group in malathion is sensitive to the choice of the gas-phase model for optimization. The solvation free energies predicted for cup08036 with the M06-2X/6-31G(d) geometry are in substantially better agreement with the experimental target (−8.2 kcal mol^{-1}, Table 1) than are those predicted for the Guthrie geometry, suggesting that similar improvements might be expected with M06-2X/6-31G(d) geometries for the several other molecules in the test set containing phosphorus ester functionality (see above). To complete this analysis, we further allowed the geometries to relax in aqueous solution at the SM8/M06-2X/6-31G(d) level. The solvation free energies for the aqueous optimized geometries were about 0.7 kcal mol^{-1} more negative in each of the three cases. This represents a 5−8% increase in the magnitude of the solvation free energy associated with geometry optimization in solution, which is in the range that we have previously found to be typical for neutral organic molecules in water[69] (this effect can be much larger when the solute has a soft mode that is coupled to a significant change in the solute's charge distribution,[70] but this situation does not hold for cup08032, cup08036, or cup08040).

As a final point for discussion, it is possible that the solvation free energies from the SM*x* models may be sensitive to the choice of density functional. To examine this point, we carried out SM8 calculations with the *m*PW1 functional[71] in place of the M06-2X functional. Over all 61 molecules in Table 1, the average unsigned deviation between the predictions from the two different functionals was 0.2 kcal mol^{-1}, which is certainly quite small.

Conclusion

The SM6, SM8, and SMD quantum mechanical continuum solvation models predict free energies of solvation for 56 molecules of the SAMPL1 test set with mean unsigned errors of 1.7, 2.1, and 1.9 kcal mol^{-1}, respectively, for the provided geometries. Given the complex functionality present in most of the test set molecules and the uncertainty in many of the experimental target values, this level of accuracy is reasonable, but there is likely room for improvement, particularly with respect to improving the representation of oxidized sulfur functionality in existing aqueous training sets. The continued identification of new and accurate experimental data will no doubt prove helpful in further model development.

Acknowledgment. This work was supported by the National Science Foundation (grants CHE06-10183 and CHE07-04974) and by the Office of Naval Research under grant No. 00014-05-01-0538.

References and Notes

(1) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492.
(2) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417.
(3) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187–4225.
(4) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161–2200.
(5) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.
(6) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley & Sons: Chichester, 2004.
(7) Cramer, C. J.; Truhlar, D. G. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1995; Vol. 6, p 1−72.
(8) Ziegler, T.; Autschbach, J. *Chem. Rev.* **2005**, *105*, 2695–2722.
(9) Dill, K. A.; Truskett, T. M.; Vlachy, V.; Hribar-Lee, B. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 173–199.
(10) Vitha, M.; Carr, P. W. *J. Chromatogr. A* **2006**, *1126*, 143–194.
(11) Chang, T. M.; Dang, L. X. *Chem. Rev.* **2006**, *106*, 1305–1322.
(12) Benjamin, I. *Chem. Rev.* **2006**, *106*, 1212–1233.
(13) Cruz, L.; Urbanc, B.; M, B. J.; Lazo, N. D.; Teplow, D. B.; Stanley, H. E. *Proc. Natl. Acad. Sci., USA* **2005**, *102*, 18258–18263.
(14) Gilson, M. K.; Zhou, H. X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
(15) Jorgensen, W. L.; Duffy, E. M. *Adv. Drug Deliv. Rev.* **2002**, *54*, 355–366.
(16) Raevsky, O. A. *Mini-Rev. Med. Chem.* **2004**, *4*, 1041–1052.
(17) Hellou, J.; Steller, S.; Leonard, J.; Langille, M. A.; Tremblay, D. *Marine Environ. Res.* **2005**, *59*, 101–117.
(18) Balakin, K. V.; Savchuk, N. P.; Tetko, I. V. *Curr. Med. Chem.* **2006**, *13*, 223–241.
(19) Johnson, S. R.; Zheng, W. F. *AAPS J.* **2006**, *8*, E27–E40.
(20) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. *J. Med. Chem.* **2008**, *51*, 769–779.
(21) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2011–2033.
(22) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
(23) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
(24) Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2046–2054.
(25) Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2008**, *112*, 8651–8655.
(26) Guthrie, J. P. *J. Phys. Chem. B* **2009**, *113*, 4501−4507.
(27) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305–8311.
(28) Cramer, C. J.; Truhlar, D. G. In *Trends and Perspectives in Modern Computational Science*; Maroulis, G., Simos, T. E., Eds.; Brill Academic: Amsterdam, 2006; Vol. 6, p 112−140.
(29) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760–768.
(30) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.*, in press (DOI: 10.1021/jp810292n).
(31) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133–1152.
(32) Hoijtink, G. J.; de Boer, E.; Van der Meij, P. H.; Weijland, W. P. *Recl. Trav. Chim. Pays-Bas* **1956**, *75*, 487–503.
(33) Jano, I. *CR. Acad. Sci. Paris* **1965**, *261*, 103–105.
(34) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Viellard, A., Eds.; Reidel: Dordrecht, 1980; Vol. 2, p 25−72.
(35) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164–170.
(36) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
(37) Chambers, C. C.; Giesen, D. J.; Hawkins, G. D.; Vaes, W. H. J.; Cramer, C. J.; Truhlar, D. G. In *Rational Drug Design*; Truhlar, D. G., Howe, W. J., Hopfinger, A. J., Blaney, J. M., Dammkoehler, R. A., Eds.; Springer: New York, 1999, p 51−72.
(38) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1995**, *16*, 422–440.
(39) Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 1057–1068.
(40) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117–9133.
(41) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032–3041.
(42) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 16066–16081.
(43) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 408–422.
(44) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theor. Comput.* **2008**, *4*, 877–887.

(45) Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Orozco, M.; Luque, F. J. *J. Mol. Struct. (THEOCHEM)* **2005**, *727*, 29–40.
(46) Pierotti, R. A. *J. Phys. Chem.* **1963**, *67*, 1840–1845.
(47) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717–726.
(48) Claverie, P. In *Intermolecular Interactions: From Diatomics to Biomolecules*; Pullman, B., Eds.; Wiley: Chichester, 1978, p 69−305.
(49) Bachs, M.; Luque, F. J.; Orozco, M. *J. Comput. Chem.* **1994**, *15*, 446–454.
(50) Luque, F. J.; Bachs, M.; Orozco, M. *J. Comput. Chem.* **1994**, *15*, 847–857.
(51) Orozco, M.; Bachs, M.; Luque, F. J. *J. Comput. Chem.* **1995**, *16*, 563–575.
(52) Curutchet, C.; Cramer, C. J.; Truhlar, D. G.; Ruiz-López, M. F.; Rinaldi, D.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2003**, *24*, 284–297.
(53) Shimizu, K.; Freitas, A. A.; Farah, J. P. S.; Dias, L. G. *J. Phys. Chem. A* **2005**, *109*, 11322–11327.
(54) Basilevsky, M. V.; Leontyev, I. V.; Luschekina, S. V.; Kondakova, O. A.; Sulimov, V. B. *J. Comput. Chem.* **2006**, *27*, 552–570.
(55) de O. Mendes, C. L.; da Silva, C. O.; da Silva, E. C. *J. Phys. Chem. A* **2006**, *110*, 4034–4041.
(56) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210–3221.
(57) Ben-Naim, A. *Solvation Thermodynamics*; Plenum: New York, 1987.
(58) Higashi, M.; Marenich, A. V.; Olson, R. M.; Chamberlin, A. C.; Pu, J.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Zhu, T.; Hawkins, G. D.; Chuang, Y.-Y.; Fast, P. L.; Lynch, B. J.; Liotard, D. A.; Rinaldi, D.; Gao, J.; Cramer, C. J.; Truhlar, D. G. *GAMESSPLUS*, version 2008-2; University of Minnesota: Minneapolis, 2008.
(59) Olson, R. M.; Marenich, A. V.; Chamberlin, A. C.; Kelly, C. P.; Xidos, J. D.; Li, J.; Thompson, J. D.; Hawkins, G. D.; Winget, P. D.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM*, version 2008; University of Minnesota: Minneapolis, 2008.
(60) Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Daschel, H.; Zhang, W.; Korambath, P. P.; Baker, J.; Byrd, E. F. C.; Voorhis, T. V.; Oumi, M.; Hirata, S.; Hsu, C.-P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. *J. Comput. Chem.* **2000**, *21*, 1532–1548.
(61) Marenich, A. V.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *GESOL*, version 2008; University of Minnesota: Minneapolis, 2008.
(62) Nakamura, H.; Xidos, J. D.; Chamberlin, A. C.; Kelly, C. P.; Valero, R.; Thompson, J. D.; Li, J.; Hawkins, G. D.; Zhu, T.; Lynch, B. J.; Volobuev, Y.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *HONDOPLUS*, v.5.1; University of Minnesota: Minneapolis, MN, 2007.
(63) Chamberlin, A. C.; Kelly, C. P.; Thompson, J. D.; Lynch, B. J.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Zhu, T.; Volobuev, Y.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *SMxGAUSS*, version 3.4.2; University of Minnesota: Minneapolis, 2007.
(64) *Jaguar 6.5*, Release 112; Schrödinger, Inc.: Portland, OR, 2005.
(65) Fendinger, N. J.; Glotfelty, D. E. *Environ. Sci. Technol.* **1988**, *22*, 1289–1293.
(66) Wauchope, R. D.; Buttler, T. M.; Hornsby, A. G.; Augustijn-Beckers, P. W. M.; Burt, J. P. In *Reviews of Environmental Contamination and Toxicology*; Ware, G. W., Ed.; Springer Verlag: New York, 1992; Vol. 123.
(67) Cramer, C. J.; Truhlar, D. G. In *Free Energy Calculations in Rational Drug Design*; Reddy, M. R., Erion, M. D., Eds.; Kluwer Academic/Plenum: New York, 2001, p 63−95.
(68) Schmuckler, M. E.; Barefoot, A. C.; Kleier, D. A.; Cobranchi, D. P. *Pest Manage. Sci.* **2000**, *56*, 521–532.
(69) Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aid. Mol. Des.* **1992**, *6*, 629–666.
(70) Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertrán, J., Eds.; Kluwer: Dordrecht, 1996, p 1−81.
(71) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664–675.

JP809094Y