Chapter 13

Universal Solvation Models

Gregory D. Hawkins[1], Tianhai Zhu[1], Jiabo Li[1], Candee C. Chambers[1,3],
David J. Giesen[1,4], Daniel A. Liotard[2], Christopher J. Cramer[1],
and Donald G. Truhlar[1]

[1]Department of Chemistry and Supercomputer Institute, University of Minnesota,
Minneapolis, MN 55455
[2]Laboratoire de Physico-Chimie Theorique, Universite de Bordeaux 1, 351 Cours
de la Liberation, 33405 Talence Cedex, France

This chapter presents an overview of the SM5 suite of universal solvation models for computing free energies of solvation in water and nonaqueous solvents. After a general review of the theoretical components of all the SM5 solvation models, we specifically compare the performance of those that have been parameterized for both aqueous and organic solvents. These are called the universal solvation models, and they include models based on semiempirical neglect of diatomic differential overlap molecular orbital theory, density functional theory, and *ab initio* Hartree-Fock theory, and also a model with implicit electrostatics.

The combination of quantum mechanics (QM) for a solute and classical mechanics (CM) for a solvent can be achieved in two main ways. In the first (*1,2*), the solvent is represented atomistically, solvent-solvent interactions are represented by an analytical force field, and solvent-solute interactions are treated by some combination of analytical force field terms (e.g., Lennard-Jones potentials) and addition of a classical potential perturbing the solute Hamiltonian and thereby the electronic structure. Analytical force fields are usually called a molecular mechanics treatment, and this kind of QM/CM treatment is called QM/MM. An alternative classical treatment of the solvent is to treat it as a continuum (*3–5*). Most continuum solvent calculations include only electrostatic interactions, in which case the solvent is treated as a dielectric medium, and the relevant classical equation is the Poisson equation or, for nonzero ionic strength, the Poisson-Boltzmann equation. However, we have proposed a series of QM/CM solvation models (*6–22*) in which, in addition to electrostatics, we include microscopic analogs of classical thermodynamic surface tension terms. These are called atomic surface tensions, and they account for short-range effects of a continuum solvent, especially those due to the first solvation shell.

Just as QM/MM calculations require a protocol for joining the QM solute and the MM solvent, the appropriate formalism for joining a solute to a dielectric continuum is the

[3]Current address: Departments of Chemistry and Physics, Mercyhurst College, 501 East 38th Street, Erie, PA 16546
[4]Eastman Kodak Company, Rochester, NY 14650

much older reaction field concept. For a polarizable solute this requires self-consistent reaction field (SCRF) theory (3–5). The formalism for a continuum treatment of first-solvation-shell effects is the concept of solvent-accessible surface area (SASA) (23,24). When a realistic solvent radius is employed to calculate the SASA, it equals the area of a surface through the center of the first solvation shell, and thus—in a continuum sense—it is proportional to the "average" number of solvent molecules in this shell.

The explicit treatment of solvent molecules in QM/MM methods leads to three significant difficulties. First is the fact that conventional MM methods (25–29) do not account for solvent electronic polarizability. Second is the slow convergence of solute-solvent interactions with respect to distance (30–32); this requires careful convergence of sums of long-range Coulomb forces. Third is the necessity for Monte Carlo or molecular dynamics averaging over the myriad of configurations that the explicit solvent molecules may assume with respect to the solute and each other. None of these issues is a difficulty for continuum models. First of all solvent bulk polarizability is included through the solvent's dielectric constant, and short-range non-bulk polarizability effects are included in the atomic surface tensions. Second, long-range electrostatic effects are not truncated because they are much more easily included (10) in a continuum theory than in a many-body simulation. Third, conformational averaging is implicit in the parameters that characterize the continuum, and explicit averaging over solvent positions is not required. Of course there are also disadvantages. For example, one cannot examine details of the solvent structure, and it is less obvious how to include nonequilibrium solvation effects. However, we shall concentrate here on the calculation of free energies of solvation under equilibrium conditions, so those disadvantages will not concern us.

A critical element in modern continuum models is that the boundary between the atomic solute and the continuum solvent is not required to have a simple shape, e.g., spherical, spheroidal, or ellipsoidal, as in the early days of Onsager, Kirkwood, and the 1970s and early 1980s (33–36). Rather the shape of the solute is determined by a superposition of partially overlapping atomic spheres, as in space filling models. A critical element in all our solvation models to date is that we use the generalized Born approximation (GBA) to the Poisson equation (6,37-42) for the electrostatic SCRF term. The solute shape enters this term through a dielectric descreening algorithm based on the work of Still et al. (4). We have improved this algorithm numerically (10), but the physical idea is unchanged. In particular, electric polarization of the solvent screens the solute's intramolecular interactions because those field lines pass through the dielectric medium representing the volume occupied by solvent (43,44). However, the interactions in one part of the solute may be descreened from the solvent by another part of the solute; thus the solvent screening and solute descreening depend on solute shape. We take account of this in one of two ways. The original approach (6,10,42) is to integrate over the electric polarization free energy density in each volume element of space occupied by solvent. This is called volume descreening. The second approach (13,15,44) is to reduce the descreening effect of solute atom A on solute atom B to an effective pair potential; this is called pairwise descreening (PD).

We originally developed a series of solvation models labeled variously SM1 through SM4 (6–13). These models combined the GBA with volume descreening for the SCRF term with the AM1 (45–48) and PM3 (49,50) general parameterizations of neglect-of-diatomic-differential-overlap (NDDO) (51) semiempirical molecular orbital theory and used an evolving series of functional forms for atomic surface tensions. The atomic surface tensions were written in terms of surface tension coefficients whose values were fit to training sets of experimental free energies of solvation. In some cases, the atomic surface tensions were functions of bond orders computed from the SCRF electronic density. Models SM1–3.1 (6–10,13) were parameterized for aqueous solution, and model SM4 (11,12) was parameterized for alkane solvents. In order to make the electrostatic portions of the models more realistic, we also developed a class IV charge model called Charge Model 1 (CM1) (52), and this charge model was used to create the reaction field in

the SM4 models. Solvation models SM1–SM4 were successful in their own right, especially SM2–SM4. In addition they provided insight into the theoretical underpinnings required for a successful solvation model.

With this experience as background, we embarked on a project to create universal solvation models, by which we mean solvation models that may be applied consistently in water and also in any organic solvent. Furthermore, motivated by the continuing usefulness of a variety of "levels" in gas-phase electronic structure theory, e.g., the NDDO level, the *ab initio* Hartree-Fock (HF) level (*53*), and density functional theory (DFT) (*54*), and motivated by the continuing usefulness of a variety of basis sets (*55*) and charge models (*52,56–60*), we have created several levels of universal solvation models based on different charge models and descreening algorithms and parameterized them for a variety of solute-electronic-structure levels and basis sets. In particular we have created universal solvation models with class IV charges, class II charges, and implicit charges and with volume and pairwise descreening, and we parameterized them for the MNDO (*61*) and MNDO/d (*62–64*) general parameterizations of NDDO theory as well as AM1 and PM3 and also for *ab initio* HF and DFT. For HF and DFT, the surface tension coefficients depend on basis set, and we have begun to find the parameters for several useful basis sets. (The universal models, i.e., those parameterized for both water and organic solvents, all use volume descreening; models based on pairwise descreening (*15*) have been parameterized only for aqueous solution.) Finally we note we have distinguished two modes of treating the solute geometry: the "flexible" mode in which the solute geometry is re-optimized in solution and the "rigid" mode in which it is not; in the latter (*19–22*) the effect of solute geometry relaxation in the vicinity of equilibrium structures is included implicitly in the parameters. The purpose of this chapter is to summarize and compare these universal solvation models (*14–22*), which are called as a group the SM5 suite of solvation models.

2. Overview of SM5 Solvation Models

The distinguishing characteristic of SM5 solvation models is the set of functional forms used for the atomic surface tensions. They depend on solute bond distances, in contrast to the earlier functional forms that depended on solute bond orders. They also depend on a small set of solvent descriptors. All SM5 solvation models use the same basic approach to atomic surface tensions, although there has been some evolution from the earlier SM5 models (*14–19*) to the later ones (*20–22*) in which these forms have become finalized.

A great deal of the effort involved in creating the SM5 suite of solvation models has been the development of a broad training set of experimental data. The training set includes a wide representation of solute functional groups and solvents, and the solutes selected for the training set are essentially free of complicating conformational issues. Furthermore we have made an attempt to eliminate incorrect and unreliable experimental data. The training set has evolved (*6–22*); in the present article all statistical characterization of SM5 models will be carried out using the final SM5 training set, and the reader should realize that this data set is larger than the training set actually used to parameterize the earlier (*14–19*) models. Although the sources of experimental data for the final SM5 training set are diverse (full references are given in our original papers; see especially Ref. (*20*)), we note that five sources of data proved especially useful (*65–69*).

A critical component in our ability to parameterize the models for general organic solvents is the use of a minimal number of descriptors to specify a solvent. In the final version of the SM5 models (*20–22*), the atomic surface tension coefficients depend on six solvent descriptors for organic solvents:

n the index of refraction for visible light

α Abraham's hydrogen bond acidity parameter $\Sigma \alpha_2^{H}$ (*71*)

β Abraham's hydrogen bond basicity parameter $\Sigma\beta_2^H$ (71)

γ the macroscopic surface tension at a solvent-air interface

ϕ the number of aromatic carbon atoms in the solvent divided by the total number of non-hydrogenic atoms in the solvent

ψ the number of F, Cl, and Br atoms in the solvent divided by the total number of nonhydrogenic atoms in the solvent.

In addition the SCRF term involves:

ε the dielectric constant

For aqueous solution, we again use ε in the SCRF term, but we do not require the functional dependence of the atomic surface tensions on n, α, β, γ, ϕ, and ψ to hold for water. Instead we parameterize water separately because of its unique properties. Furthermore, because the water training set is more sensitive to electrostatic effects than is the organic one, we parameterize water first in order to determine acceptable values for the intrinsic Coulomb radii (which is what we call the atomic radii used in the SCRF calculation; we also call them electrostatic radii); these intrinsic Coulomb radii are then fixed at the same values for calculations in organic solvents.

The calculations also require atomic radii for the calculation of first-solvation-shell effects. For these radii, all SM5 models use Bondi's van der Waals radii (70) without any adjustment.

Beginning with Ref. (20), all functional forms, radii, and training sets are frozen, and only the solute wave function and charge model and the surface tension coefficients change in Refs. (21) and (22).

We checked for two different SM5 models whether including functions of ε in the solvent descriptors for surface tensions would reduce the mean errors. Interestingly, it does not, at least to any meaningful extent. This may indicate that our electrostatic radii are reasonable. If they were not reasonable, then effects associated with the "first solvation shell" would be expected to have a significant electrostatic component.

3. The SM5 family

The SM5 family now consists of seven solvation models, some of which have been parameterized for more than one method for obtaining the solute wave function. The models and their distinguishing features are

SM5.42R   rigid model based on class IV charges obtained by Charge Model 2 (CM2) (60) and on volume descreening

SM5.4   flexible model based on class IV charges obtained by CM1 (52) and on volume descreening

SM5.4PD   flexible model based on class IV charges obtained by CM1 (52) and on pairwise descreening

SM5.2R   rigid model based on class II charges (56,57) and on volume descreening

SM5.2PD   flexible model based on class II charges (56) and on pairwise descreening

SM5.0R   rigid model employing only atomic surface tensions (i.e., all the charges are set equal to zero, and electrostatic effects are included only implicitly though the atomic surface tensions)

SM5.05R   like SM5.0R except that class I charges are added to some of the atoms in ions and zwitterions, and electrostatic interactions are calculated using volume descreening.

Table 1. Parameterizations of universal solvation models

model	solute wave function	Ref.
SM5.42R	BPW91/MIDI!(6D)	(21)
	BPW91/DZVP	(21)
	HF/MIDI!(6D)	(22a)
	AM1	(22b)
	PM3	(22b)
SM5.4	AM1	(14,16–18)
	PM3	(14,16–18)
SM5.2R	AM1	(20)
	PM3	(20)
	MNDO	(20)
	MNDO/d	(20)
SM5.0R	–	(19)

Note that .4, .2, or .0 denotes the choice of charge model (in .0 models, the charges are zero, or, more precisely, they are implicit in the surface tensions), R or its absence denotes rigid or flexible, and PD or its absence denotes the choice between pairwise or volume descreening. The highest possible level would be an SM5.42 flexible model with volume descreening, but we have not created such a model yet.

For levels employing class II or higher charges, the surface tension parameters are reoptimized for each level of solute wave function (because these empirical parameters can make up for any systematic deficiencies in a given choice of solute level). Table 1 gives a complete list of the solute wave functions for which each of the five levels that involve either class II and class IV charges has been optimized to date. The notation indicates the wave function after a slash (solidus). Thus SM5.42/HF/MIDI!(6D) indicates that the wave function is calculated at the HF/MIDI!(6D) level. The basis sets examined so far are MIDI!(6D) (60,72) and DZVP (73).

The rigid models are designed to be used with any reasonable gas-phase geometry, and some attempt was made to parameterize them in such a way as to not be overly sensitive to small changes in geometries (although some residual sensitivity remains, especially in SM5.42R models due to the dependence of CM2 charges on geometry). As a matter of completeness, we note that *all* rigid models are parameterized using HF/MIDI! geometries. (Note: The MIDI! basis set (72) uses five spherical harmonic d functions in each d shell, and the MIDI!(6D) basis (60, 72) differs only in that it uses six Cartesian d functions in each d shell. The results obtained with these two basis are very similar. The MIDI! basis is therefore preferred because it is smaller, but some program packages do not support spherical harmonic d functions, so the MIDI!(6D) basis set can be used with a larger number of electronic structure programs.)

4. Comparison of Universal SM5 Solvation Models

In this section we compare the universal SM5 solvation models. There are twelve universal parameterizations, and they are listed in Table 1.

In the rest of this paper, all results for flexible models are based on separate geometry optimizations in the gas phase and in solution, and all results for rigid models are based on HF/MIDI! geometries, with one exception. The exception is the SM5.0R model, where results are shown for HF/MIDI! gas-phase geometries, AM1 gas-phase geometries, and MM3* gas-phase geometries, where MM3* is the MACROMODEL (74) implementation of the MM3 (29,75–77) force field. Ordinarily these would be denoted SM5.0R//HF/MIDI!, SM5.0R//AM1, and SM5.0R//MM3*. In the present paper, to

Table 2. Mean Unsigned Error (kcal/mol) of Selected SM*x* Models by Solvent Functional Group Class

| Solvent Class | Number of | | | SM5.42R/BPW91/... | | | SM5.4/... | |
	Solvents[a]	Solute classes[b]	Data[c]	DZVP	MIDI!(6D)	SM5.42R/HF/ MIDI!(6D)	AM1	PM3
Aqueous	1	31	248	0.41	0.43	0.46	0.55	0.49
Alkanes	11	30	475	0.32	0.33	0.32	0.33	0.31
Cycloalkanes	2	24	106	0.38	0.38	0.40	0.41	0.39
Arenes	12	16	256	0.38	0.38	0.39	0.42	0.38
Aliphatic alcohols	12	31	299	0.57	0.55	0.52	0.57	0.53
Aromatic alcohols	2	7	12	0.59	0.55	0.53	0.71	0.68
Ketones	4	10	35	0.43	0.49	0.45	0.48	0.40
Esters	2	8	36	0.53	0.60	0.54	0.44	0.42
Aliphatic ethers	4	19	99	0.49	0.52	0.52	0.57	0.50
Aromatic ethers	3	5	15	0.53	0.59	0.53	0.32	0.30
Amines	2	6	12	0.32	0.36	0.37	0.66	0.56
Pyridines	3	5	15	0.46	0.49	0.40	0.33	0.34
Nitriles	2	5	10	0.69	0.76	0.59	0.54	0.46
Nitro compounds	4	8	27	0.56	0.63	0.58	0.68	0.60
Tertiary amides	2	5	10	0.37	0.38	0.26	0.33	0.31
Haloaliphatics	12	27	269	0.46	0.47	0.47	0.58	0.48
Haloaromatics	6	11	106	0.44	0.40	0.38	0.50	0.45
Miscellaneous acidic solvents	3	5	15	0.49	0.48	0.40	0.41	0.37
Miscellaneous non-acidic solvents	4	12	39	0.41	0.42	0.43	0.37	0.35
Total:	91	31	2084	0.43	0.43	0.43	0.48	0.43

[a] Number of solvents in this solvent class

[b] Number of solutes classes for which data exists in this solvent class. The solute classes are listed in Tables 6–9.

[a] Total number of solute/solvent data involving this solvent class

Table 3. Mean Unsigned Error (kcal/mol) of Selected SMx Models by Solvent Functional Group Class

Solvent Class	SM5.42R/...		SM5.2R/...				SM5.0R//...	
	AM1	PM3	MNDO/d	MNDO	AM1	PM3	HF/MIDI!	AM1
Aqueous	0.42	0.43	0.47	0.47	0.44	0.51	0.53	0.57
Alkanes	0.31	0.32	0.29	0.29	0.32	0.30	0.29	0.31
Cycloalkanes	0.39	0.39	0.36	0.36	0.37	0.35	0.35	0.36
Arenes	0.40	0.39	0.30	0.30	0.34	0.33	0.28	0.29
Aliphatic alcohols	0.53	0.54	0.46	0.47	0.53	0.49	0.44	0.45
Aromatic alcohols	0.54	0.54	0.49	0.49	0.52	0.51	0.48	0.48
Ketones	0.53	0.53	0.35	0.35	0.42	0.40	0.42	0.41
Esters	0.52	0.54	0.50	0.50	0.59	0.56	0.51	0.53
Aliphatic ethers	0.52	0.53	0.46	0.46	0.50	0.47	0.44	0.43
Aromatic ethers	0.52	0.54	0.43	0.43	0.49	0.46	0.37	0.34
Amines	0.39	0.41	0.45	0.45	0.43	0.40	0.55	0.55
Pyridines	0.42	0.42	0.30	0.30	0.43	0.38	0.32	0.35
Nitriles	0.76	0.73	0.45	0.45	0.60	0.56	0.50	0.47
Nitro compounds	0.58	0.64	0.55	0.55	0.47	0.55	0.47	0.49
Tertiary amides	0.34	0.32	0.23	0.23	0.36	0.34	0.43	0.47
Haloaliphatics	0.48	0.48	0.45	0.45	0.45	0.45	0.53	0.51
Haloaromatics	0.54	0.51	0.27	0.27	0.34	0.29	0.29	0.28
Miscellaneous acidic solvents	0.48	0.47	0.38	0.38	0.52	0.48	0.31	0.36
Miscellaneous non–acidic solvents	0.39	0.40	0.42	0.42	0.46	0.44	0.58	0.59
Total:	0.43	0.44	0.38	0.39	0.41	0.40	0.40	0.41

Table 4. Mean Signed Error (kcal/mol) of Selected SMx Models by Solvent Functional Group Class

| Solvent Class | Number of | | | SM5.42R/BPW91/... | | SM5.42R/HF/ | SM5.4/... | |
	Solvents[a]	Solute classes[b]	Data[c]	DZVP	MIDI!(6D)	MIDI!(6D)	AM1	PM3
Aqueous	1	31	248	-0.01	-0.03	-0.03	0.04	0.04
Alkanes	11	30	475	-0.03	-0.06	-0.04	-0.10	-0.11
Cycloalkanes	2	24	106	0.14	0.07	0.01	-0.02	-0.06
Arenes	12	16	256	0.25	0.19	0.24	0.20	0.19
Aliphatic alcohols	12	31	299	0.00	-0.05	-0.08	-0.11	-0.12
Aromatic alcohols	2	7	12	0.25	0.07	0.07	0.51	0.44
Ketones	4	10	35	-0.14	-0.20	-0.15	-0.35	-0.17
Esters	2	8	36	0.41	0.37	0.36	0.27	0.29
Aliphatic ethers	4	19	99	0.07	-0.02	-0.05	0.02	0.07
Aromatic ethers	3	5	15	-0.44	-0.58	-0.51	-0.02	-0.01
Amines	2	6	12	0.19	0.03	0.01	0.56	0.46
Pyridines	3	5	15	-0.17	-0.31	-0.24	0.10	0.13
Nitriles	2	5	10	-0.66	-0.76	-0.59	-0.54	-0.41
Nitro compounds	4	8	27	-0.36	-0.30	-0.22	-0.65	-0.52
Tertiary amides	2	5	10	-0.03	-0.14	-0.04	-0.12	-0.08
Haloaliphatics	12	27	269	-0.10	-0.12	-0.11	0.01	-0.01
Haloaromatics	6	11	106	-0.34	-0.29	-0.24	-0.30	-0.23
Miscellaneous acidic solvents	3	5	15	0.16	0.02	-0.05	0.02	-0.04
Miscellaneous non-acidic solvents	4	12	39	-0.08	-0.10	-0.08	-0.07	-0.13
Total:	91	31	2084	0.00	-0.04	-0.03	-0.03	-0.03

[a] Number of solvents in this solvent class

[b] Number of solutes classes for which data exists in this solvent class. The solute classes are listed in Tables 6–9.

[c] Total number of solute/solvent data involving this solvent class

Table 5. Mean Signed Error (kcal/mol) of Selected SM*x* Models by Solvent Functional Group Class

Solvent Class	SM5.42R/...		SM5.2R/...				SM5.0R//...	
	AM1	PM3	MNDO/d	MNDO	AM1	PM3	HF/MIDI!	AM1
Aqueous	-0.04	-0.04	–0.05	–0.07	–0.03	–0.01	0.06	0.11
Alkanes	-0.03	-0.04	–0.05	–0.05	–0.06	–0.05	-0.05	-0.06
Cycloalkanes	0.13	0.12	0.00	0.00	0.07	0.04	-0.07	-0.08
Arenes	0.27	0.27	0.08	0.08	0.16	0.13	-0.05	-0.07
Aliphatic alcohols	-0.06	-0.06	–0.07	–0.09	–0.03	–0.04	-0.03	0.00
Aromatic alcohols	0.17	0.15	0.16	0.16	0.16	0.16	0.14	0.16
Ketones	-0.27	-0.24	0.02	0.02	–0.07	–0.02	0.25	0.27
Esters	0.36	0.35	0.42	0.42	0.50	0.46	0.43	0.46
Aliphatic ethers	-0.02	-0.01	0.02	0.02	0.08	0.07	0.09	0.11
Aromatic ethers	-0.48	-0.49	–0.40	–0.40	–0.43	–0.43	-0.33	-0.29
Amines	0.22	0.19	–0.04	–0.04	0.10	0.03	-0.21	-0.18
Pyridines	-0.24	-0.23	–0.09	–0.09	–0.13	–0.11	0.03	0.10
Nitriles	-0.76	-0.73	–0.27	–0.27	–0.53	–0.40	0.06	0.11
Nitro compounds	-0.41	-0.37	–0.02	–0.01	–0.23	–0.14	0.25	0.24
Tertiary amides	-0.10	-0.07	0.13	0.13	0.03	0.09	0.28	0.37
Haloaliphatics	-0.16	-0.15	–0.05	–0.05	–0.10	–0.07	0.04	0.02
Haloaromatics	-0.43	-0.41	–0.06	–0.06	–0.22	–0.15	0.14	0.15
Miscellaneous acidic solvents	0.19	0.19	0.16	0.16	0.17	0.18	0.10	0.15
Miscellaneous non–acidic solvents	-0.05	-0.05	–0.13	–0.13	–0.21	–0.18	-0.21	-0.15
Total:	-0.03	-0.03	–0.02	–0.03	–0.02	–0.01	0.01	0.02

Table 6. Mean Unsigned Error of Selected SMx Models by Solute Functional Group Class

	Number of			SM5.42R/BPW91/...		SM5.42R/HF/	SM5.4/...	
Solute Class	Solutes[a]	Solvent classes[b]	Data[c]	DZVP	MIDI!(6D)	MIDI!(6D)	AM1	PM3
Unbranched alkanes	9	19	84	0.45	0.40	0.38	0.42	0.44
Branched alkanes	5	3	12	0.40	0.42	0.45	0.50	0.51
Cycloalkanes	5	6	18	0.35	0.52	0.43	0.81	0.83
Alkenes	9	4	27	0.33	0.28	0.25	0.48	0.35
Alkynes	5	3	14	0.26	0.21	0.16	0.26	0.21
Arenes	9	19	134	0.50	0.57	0.51	0.31	0.29
Alcohols	17	19	385	0.37	0.37	0.36	0.38	0.36
Ethers	12	19	93	0.42	0.48	0.44	0.47	0.43
Aldehydes	7	8	38	0.55	0.41	0.41	0.44	0.45
Ketones	12	18	203	0.44	0.44	0.40	0.42	0.42
Carboxylic acids	5	14	124	0.41	0.45	0.43	0.64	0.62
Esters	14	8	249	0.33	0.33	0.34	0.49	0.45
Bifunctional compounds containing H, C, O	5	8	28	1.16	1.07	1.11	0.87	0.86
Inorganic compounds containing H and O	2	9	22	0.60	0.57	0.63	1.28	1.06
Aliphatic Amines	15	10	168	0.36	0.34	0.31	0.40	0.34
Aromatic Amines	11	12	81	0.30	0.33	0.35	0.46	0.44
Nitriles	4	6	22	0.48	0.44	0.90	0.42	0.38
Nitrohydrocarbons	6	8	38	0.55	0.43	0.40	0.53	0.23
Amides & ureas	4	6	11	1.08	1.24	1.29	1.93	0.93
Bifunctional compounds containing N	6	3	11	0.61	0.69	0.70	1.76	1.03
Inorganic compounds containing N	2	8	15	0.64	0.70	0.69	1.92	0.58
Thiols	4	5	14	0.31	0.33	0.42	0.34	0.29
Sulfides	6	6	23	0.61	0.59	0.68	0.72	0.68
Disulfides	2	3	5	0.23	0.25	0.19	0.10	0.10
Fluorinated hydrocarbons	9	5	19	0.35	0.49	0.54	0.49	0.33
Chloroalkanes	13	5	35	0.49	0.34	0.32	0.29	0.27
Chloroalkenes	5	4	16	0.65	0.65	0.61	0.46	0.38
Chloroarenes	8	6	37	0.33	0.43	0.61	0.32	0.28
Brominated hydrocarbons	14	6	50	0.34	0.36	0.35	0.25	0.28
Iodinated hydrocarbons	9	6	28	0.18	0.34	0.31	0.18	0.20
Multifunctional halogenated solutes	26	9	80	0.75	0.70	0.67	0.62	0.67
Total:	260	19	2084	0.43	0.43	0.43	0.48	0.43

[a]Number of solutes in this solute class [b]Number of solvent classes for which there are data for this solute class. The solvent classes are listed in Tables 2–5. [c]Total number of solute/solvent data involving solutes in this solute class

Table 7. Mean Unsigned Error (kcal/mol) of Selected SM*x* Models by Solute Functional Group Class

Solute Class	SM5.42R/...		SM5.2R/...				SM5.0R//...	
	AM1	PM3	MNDO/d	MNDO	AM1	PM3	HF/MIDI!	AM1
Unbranched alkanes	0.40	0.41	0.38	0.38	0.45	0.42	0.41	0.43
Branched alkanes	0.41	0.41	0.45	0.45	0.39	0.42	0.41	0.44
Cycloalkanes	0.33	0.31	0.34	0.34	0.31	0.30	0.38	0.37
Alkenes	0.31	0.28	0.21	0.21	0.31	0.25	0.20	0.19
Alkynes	0.15	0.14	0.13	0.13	0.18	0.17	0.11	0.15
Arenes	0.50	0.53	0.39	0.39	0.53	0.47	0.39	0.38
Alcohols	0.37	0.38	0.33	0.33	0.35	0.34	0.35	0.36
Ethers	0.51	0.49	0.39	0.39	0.40	0.39	0.43	0.45
Aldehydes	0.47	0.54	0.46	0.46	0.41	0.43	0.51	0.54
Ketones	0.42	0.44	0.33	0.33	0.34	0.35	0.36	0.37
Carboxylic acids	0.46	0.47	0.35	0.35	0.39	0.38	0.40	0.40
Esters	0.40	0.39	0.29	0.29	0.33	0.31	0.28	0.31
H, C, O Bifunctional compounds	1.00	1.00	0.86	0.86	1.07	0.99	0.85	0.83
H and O Inorganic compounds	0.68	0.66	0.58	0.58	0.57	0.58	0.65	0.65
Aliphatic Amines	0.34	0.31	0.27	0.27	0.32	0.30	0.26	0.32
Aromatic Amines	0.35	0.33	0.43	0.43	0.35	0.39	0.50	0.52
Nitriles	0.54	0.54	0.35	0.35	0.39	0.26	0.51	0.49
Nitrohydrocarbons	0.24	0.29	0.23	0.23	0.57	0.40	0.24	0.25
Amides & ureas	1.08	1.01	1.49	1.49	1.48	1.32	1.86	1.89
N Bifunctional compounds	0.66	0.64	0.96	0.96	0.88	0.91	1.02	1.02
N Inorganic compounds	0.69	0.57	0.57	0.57	0.62	0.53	0.71	0.73
Thiols	0.32	0.34	0.48	0.45	0.43	0.40	0.38	0.43
Sulfides	0.53	0.67	0.98	0.84	0.90	0.83	0.66	0.67
Disulfides	0.26	0.22	0.16	0.17	0.16	0.17	0.19	0.27
Fluorinated hydrocarbons	0.46	0.50	0.67	0.68	0.42	0.46	0.70	0.69
Chloroalkanes	0.40	0.48	0.27	0.36	0.37	0.51	0.38	0.41
Chloroalkenes	0.64	0.58	0.65	0.54	0.73	0.88	0.78	0.82
Chloroarenes	0.45	0.39	0.56	0.64	0.33	0.37	0.29	0.29
Brominated hydrocarbons	0.37	0.44	0.33	0.35	0.42	0.33	0.37	0.37
Iodinated hydrocarbons	0.26	0.37	0.41	0.48	0.49	0.37	0.43	0.41
Multifunctional halogenated solutes	0.71	0.72	0.64	0.69	0.58	0.79	0.68	0.68
Total:	0.43	0.44	0.38	0.39	0.41	0.40	0.40	0.41

Table 8. Mean Signed Error of Selected SM*x* Models by Solute Functional Group Class

Solute Class	Number of Solutes[a]	Solvent classes[b]	Data[c]	SM5.42R/BPW91/... DZVP	SM5.42R/BPW91/... MIDI!(6D)	SM5.42R/HF/ MIDI!(6D)	SM5.4/... AM1	SM5.4/... PM3
Unbranched alkanes	9	19	84	0.21	0.09	0.10	0.13	0.17
Branched alkanes	5	3	12	0.08	0.13	0.18	-0.08	-0.10
Cycloalkanes	5	6	18	-0.31	-0.47	-0.32	0.71	0.77
Alkenes	9	4	27	0.19	-0.02	-0.04	0.39	0.27
Alkynes	5	3	14	-0.03	-0.01	0.02	-0.05	-0.04
Arenes	9	19	134	-0.41	-0.53	-0.47	-0.10	-0.12
Alcohols	17	19	385	-0.01	0.01	0.02	0.04	0.03
Ethers	12	19	93	0.11	0.04	0.10	0.02	0.11
Aldehydes	7	8	38	0.38	-0.08	-0.21	0.08	-0.03
Ketones	12	18	203	-0.25	-0.28	-0.19	-0.07	-0.07
Carboxylic acids	5	14	124	-0.05	-0.20	-0.11	0.30	0.26
Esters	14	8	249	0.08	0.13	0.10	-0.30	-0.27
Bifunctional compounds containing H, C, O	5	8	28	0.94	0.83	0.75	0.38	0.33
Inorganic compounds containing H and O	2	9	22	0.00	-0.01	0.03	-1.27	-1.05
Aliphatic amines	15	10	168	0.10	0.08	0.07	-0.10	0.00
Aromatic amines	11	12	81	0.00	-0.01	0.07	-0.33	-0.18
Nitriles	4	6	22	0.01	0.00	-0.90	-0.03	0.02
Nitrohydrocarbons	6	8	38	0.03	0.01	0.01	-0.05	-0.01
Amides & ureas	4	6	11	0.73	0.86	0.83	1.01	0.29
Bifunctional compounds containing N	6	3	11	-0.43	-0.43	-0.23	0.62	-0.69
Inorganic compounds containing N	2	8	15	-0.35	-0.42	-0.41	0.14	0.46
Thiols	4	5	14	0.24	0.31	0.39	0.05	0.08
Sulfides	6	6	23	-0.09	-0.16	-0.14	-0.43	-0.43
Disulfides	2	3	5	0.00	-0.01	0.00	-0.02	0.10
Fluorinated hydrocarbons	9	5	19	-0.22	-0.36	-0.42	0.45	0.26
Chloroalkanes	13	5	35	-0.45	-0.09	0.11	-0.19	-0.18
Chloroalkenes	5	4	16	0.65	0.65	0.62	0.45	0.37
Chloroarenes	8	6	37	0.11	-0.41	-0.59	-0.10	-0.13
Brominated hydrocarbons	14	6	50	-0.20	-0.27	-0.23	-0.10	-0.20
Iodinated hydrocarbons	9	6	28	0.01	-0.05	-0.02	-0.03	-0.04
Multifunctional halogenated solutes	26	9	80	0.30	0.29	0.21	0.03	-0.03
Total:	260	19	2084	0.00	-0.04	-0.03	-0.03	-0.03

[a] Number of solutes in this solute class [b] Number of solvent classes for which there are data for this solute class. The solvent classes are listed in Tables 2–5. [c] Total number of solute/solvent data involving solutes in this solute class

Table 9.   Mean Signed Error (kcal/mol) of Selected SM*x* Models by Solute Functional Group Class

Solute Class	SM5.42R/...		SM5.2R/...				SM5.0R//...	
	AM1	PM3	MNDO/d	MNDO	AM1	PM3	HF/MIDI!	AM1
Unbranched alkanes	0.10	0.14	0.05	0.05	0.16	0.16	0.06	0.18
Branched alkanes	0.08	0.13	0.07	0.07	0.01	0.08	0.21	0.30
Cycloalkanes	-0.27	-0.19	–0.14	–0.14	–0.25	–0.12	-0.16	-0.07
Alkenes	0.20	0.17	–0.05	–0.05	0.21	0.11	-0.05	0.03
Alkynes	-0.01	-0.01	0.04	0.04	–0.01	0.00	0.04	0.09
Arenes	-0.44	-0.49	–0.27	–0.27	–0.42	–0.39	-0.13	-0.07
Alcohols	-0.04	-0.03	–0.01	–0.01	–0.02	–0.01	-0.01	0.02
Ethers	0.16	0.18	0.09	0.09	0.09	0.11	0.12	0.16
Aldehydes	-0.39	-0.49	–0.27	–0.27	–0.15	–0.23	-0.14	-0.21
Ketones	-0.13	-0.14	–0.09	–0.09	–0.18	–0.15	-0.07	-0.14
Carboxylic acids	-0.03	-0.04	–0.04	–0.04	–0.06	–0.06	0.05	-0.06
Esters	0.03	0.04	0.02	0.02	0.04	0.04	-0.04	-0.16
H, C, O Bifunctional compounds	0.83	0.81	0.41	0.41	0.76	0.62	0.24	0.19
H and O Inorganic compounds	0.00	0.00	0.00	0.00	–0.02	0.00	-0.08	-0.09
Aliphatic amines	0.06	0.05	0.01	0.01	0.07	0.02	0.00	0.14
Aromatic amines	0.04	0.00	0.14	0.14	–0.02	0.14	0.18	0.25
Nitriles	0.00	0.00	0.09	0.09	0.32	–0.09	0.11	0.12
Nitrohydrocarbons	0.00	0.00	0.00	0.00	0.02	–0.06	-0.02	-0.05
Amides & ureas	0.64	0.63	1.38	1.38	1.38	1.26	1.51	1.49
N Bifunctional compounds	-0.36	-0.27	–0.89	–0.89	–0.83	–0.84	-0.96	-0.82
N Inorganic compounds	-0.40	-0.24	–0.18	–0.18	–0.28	–0.08	-0.15	-0.18
Thiols	0.28	0.33	0.34	0.30	0.31	0.29	0.28	0.37
Sulfides	-0.08	-0.01	–0.13	–0.10	–0.12	–0.06	-0.24	-0.10
Disulfides	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18
Fluorinated hydrocarbons	-0.26	-0.29	–0.41	–0.44	–0.11	–0.10	0.14	0.23
Chloroalkanes	-0.34	-0.39	–0.01	–0.16	–0.27	0.43	0.09	0.03
Chloroalkenes	0.64	0.58	0.65	0.52	0.73	0.88	0.78	0.82
Chloroarenes	-0.40	-0.33	–0.54	–0.62	–0.18	–0.32	-0.08	-0.02
Brominated hydrocarbons	-0.31	-0.39	–0.19	–0.26	–0.10	–0.21	0.05	0.05
Iodinated hydrocarbons	-0.05	-0.09	–0.06	–0.04	–0.03	–0.09	-0.02	0.06
Multifunctional halogenated solutes	0.28	0.31	0.08	0.15	0.10	0.03	0.07	0.09
Total:	-0.03	-0.03	–0.02	–0.03	–0.02	–0.01	0.01	0.02

Table 10. Application of the SM5.0R model to molecular mechanics (MM3*) geometries in organic solvents. (Note nitrohydrocarbons and hydrogen were left out because not enough types in MM3* to run these molecules.)

Solute Class	Number of			SM5.4/AM1		SM5.0R		SM5.0R //MM3*	
	Solutes[a]	Classes[b]	Data[c]	MSE[d]	MUE[e]	MSE[d]	MUE[e]	MSE[d]	MUE[e]
Unbranched alkanes	9	18	76	0.21	0.40	0.12	0.40	0.22	0.43
Branched alkanes	5	2	7	0.35	0.37	0.49	0.52	0.56	0.57
Cycloalkanes	4	5	13	1.04	1.04	-0.06	0.35	0.01	0.35
Alkenes	8	3	18	0.42	0.49	-0.04	0.19	0.02	0.21
Alkynes	5	2	9	-0.07	0.27	0.02	0.12	0.08	0.14
Arenes	9	18	126	-0.11	0.32	-0.17	0.38	-0.11	0.36
Alcohols	17	18	369	0.03	0.38	-0.02	0.35	-0.02	0.35
Ethers	12	18	81	-0.04	0.43	0.12	0.41	0.13	0.41
Aldehydes	7	7	32	0.15	0.46	-0.12	0.54	-0.13	0.54
Ketones	12	17	191	-0.09	0.42	-0.10	0.36	-0.09	0.35
Carboxylic acids	5	13	119	0.28	0.64	0.03	0.39	0.03	0.39
Esters	13	7	236	-0.30	0.49	-0.02	0.28	-0.03	0.29
Bifunctional compounds containing H, C, O	4	7	23	0.44	0.98	0.41	0.92	0.41	0.92
Inorganic compounds containing H and O	1	8	18	-1.38	1.38	0.01	0.68	0.01	0.68
Aliphatic amines	11	9	153	-0.12	0.36	0.00	0.24	0.09	0.26
Aromatic amines	11	11	71	-0.29	0.42	0.16	0.43	0.23	0.44
Nitriles	4	5	18	-0.11	0.40	0.14	0.47	0.17	0.47
Amides & Ureas	2	5	7	1.01	1.93	1.32	1.66	1.38	1.73
Bifunctional compounds containing N	4	2	6	1.46	2.49	-0.88	0.97	-0.88	0.96
Inorganic compounds containing N	2	7	13	0.10	1.79	-0.21	0.66	-0.24	0.66
Thiols	3	4	10	0.07	0.35	0.37	0.40	0.40	0.42
Sulfides	6	5	17	-0.51	0.76	-0.21	0.71	-0.18	0.70
Disulfides	2	2	3	-0.04	0.04	0.00	0.27	0.10	0.28
Fluorinated hydrocarbons	5	4	13	0.39	0.45	0.02	0.51	0.07	0.51
Chloroalkanes	7	4	22	-0.21	0.30	0.05	0.35	0.06	0.37
Chloroalkenes	4	3	11	0.36	0.37	0.70	0.70	0.74	0.74
Chloroarenes	6	5	29	-0.11	0.35	-0.15	0.30	-0.10	0.29
Brominated hydrocarbons	14	5	36	-0.11	0.25	-0.02	0.36	0.01	0.36
Iodinated hydrocarbons	9	5	20	-0.01	0.15	-0.02	0.50	0.03	0.49
Multifunctional halogenated solutes	19	8	55	-0.01	0.61	0.13	0.64	0.13	0.63
Total:	220		1802	-0.03	0.48	0.00	0.39	0.05	0.39

[a]Number of solutes in this solute class [b]Number of solvent classes for which there are data for this solute class [c]Total number of solute/solvent data involving solutes in this solute class [d]Mean signed error over data in solute class [e]Mean unsigned error over solute class

simplify the discussion and tables, we will use the convention that //HF/MIDI! is implied for rigid models when no geometry is indicated.

All SM5 models are parameterized for solutes containing H, C, N, O, F, S, Cl, Br, and I, and—in addition—some of them are parameterized for solutes containing P. The final SM5 training set has 2135 solvation free energies for 275 neutral solutes in 91 solvents and 43 free energies of solvation for ions in water. In this section we will not discuss phosphorus-containing compounds or ions. For neutral solutes without phosphorus, the final SM5 training set has 2084 solvation free energies for 260 solutes in 91 solvents. The non-phosphorus solutes are classified into 31 classes and the solvents into 19.

Tables 2 and 3 show mean unsigned errors for each of the solvent classes. The bottom row gives the overall mean unsigned error, which ranges from 0.38 to 0.48 kcal/mol. The models with class II charges and those lacking charges altogether perform better than the models with class IV charges. Nevertheless, based on our extensive validation for gas-phase dipole moments (60), we believe that the models with class IV charges are more physical and will be more reliable for solute functionalities that are not represented in the training set and transition states. The robust nature of the SM5 suite is indicated by the remarkably even spread of errors for individual solvent classes. Tables 4 and 5 show the breakdown by solvent class of the mean *signed* errors. Here there is more variation in the errors, with free energies of solvation systematically overestimated or underestimated for certain solvent classes. The most poorly treated solvent classes are nitriles, nitro compounds, and haloaromatics. Nevertheless, the results are quite encouraging; the median absolute value of the solvent-class mean signed errors in Tables 4 and 5 is only 0.12 kcal/mol, and the mean absolute value of these 247 mean signed errors is only 0.17 kcal/mol. The SM5.2R models, especially those based on MNDO and MNDO/d, do the best at minimizing these systematic errors, and the SM5.42R/AM1 and SM5.42R/PM3 models do the least well.

Tables 6 and 7 show the breakdown of mean unsigned errors by solute class, and Tables 8 and 9 show the breakdown of mean signed errors by solute class. The single most significant conclusion that one can draw from these tables is that the various models and parameterizations do about equally well. This is extremely encouraging; apparently the empirical surface tensions do an excellent job of making up for any systematic deficiencies in the wave functions and partial atomic charges.

The magnitudes of the class II charges appear to be somewhat smaller, on average, than the magnitudes of class IV charges. In general, we see that there is a tendency for models that underestimate the accurate partial atomic charges or treat them implicitly to be slightly easier to parameterize for free energies of solvation. This may be due to reducing instability in the electrostatic portion of the calculation.

Table 10 is of special interest for this symposium because it demonstrates the use of a rigid SM5 model with gas-phase geometries that were calculated using the molecular mechanical method MM3*. Since the MM3* model does not have sufficient types to determine geometries for H_2 or solutes containing nitro groups the 7 such molecules in our testing set were omitted. In addition, we limited our examination to solvation free energies in organic solvents (not water), for a total of 220 neutral solutes and 1802 data points. For this set of solutes, Table 10 shows that the SM5.0R model applied with MM3* gas-phase geometries produces extremely similar results to the SM5.0R model applied to HF/MIDI! geometries. The overall mean unsigned error is 0.38 kcal/mol using either set of geometries. The mean signed error over the whole set of molecules increases only from 0.00 to 0.02 when using the MM3* geometries versus the HF/MIDI! geometries with which the SM5.0R model was parameterized. For comparison, results for the SM5.4/AM1 model, which is based on self-consistently relaxed solute geometries, are also shown.

A question often asked by referees is: how well would the model do if the test set included molecules not in the training set? To address this, we performed a series of systematic tests for the SM5.2R/MNDO/d model in which we omitted one quarter of the

16

data (over 500 data) from the training set which has 2082 data. Thus the model was trained on 1561–1562 data, but tested on the full 2082 data. The mean unsigned error increased by less than 0.01 kcal/mol, which is less than 3% of the mean unsigned error of 0.38 kcal/mol in Table 3. This is consistent with our general experience. We believe that the most robust model is obtained by parameterizing with a large, broad training set.

Although the present manuscript is primarily concerned with neutral solutes, we also tested the models for ions in aqueous solution. The SM5.2R/MNDO/d model yields a mean unsigned error of 3.8 kcal/mol for 43 ions without phosphorus, and the results for other models are similar. Considering that the experimental data are uncertain by about ±5 kcal/mol for ions (as compared to about ±0.2 kcal/mol for neutrals), this appears to be quite acceptable.

5. Platforms

The SM5.4/AM1 and /PM3 parameterizations, all the parameterizations of PD models, the SM5.0R model, and the SM5.05R model are available in the latest released version of AMSOL (78) and SM5.4R/AM1 and /PM3 and SM5.2R/AM1, /PM3, and /MNDO will be in the next version of AMSOL (79).

SM5.0R and SM5.05R are also available in OMNISOL (80).

SM5.2R/AM1, /PM3, /MNDO, and /MNDO/d have been implemented in AMPAC (81).

SM5.42R/BPW91/MIDI!(6D) and /BPW91/DZVP have been implemented in DGAUSS (82) and GAUSSIAN94 (83).

SM5.4R/HF/MIDI!(6D) has been implemented in GAUSSIAN94 (83) and GAMESS (84,85).

Additional implementations involving other density functionals and other basis sets are in progress.

5. Concluding remarks

The SM5 suite of solvation models provides reasonable accuracy for free energies of solvation over a broad range of solute functionality and solvent type. The results are just as accurate on the average (sometimes even more accurate!) when affordable low levels of electronic structure theory are used. This approach therefore provides a convenient way to mix a quantum mechanical treatment of the solute with a classical treatment of the solvent. Careful attention to the first solvation shell, which "links" the quantal and classical regions, allows one to obtain quantitative accuracy for both ions and neutrals of arbitrary shape.

6. Acknowledgments

References

(1) Bash, P. A.; Field, M. J.; Karplus, M. *J. Comp. Chem.* **1990**, *11*, 700.
(2) Gao, J. *J. Phys. Chem.* **1992**, *96*, 537.
(3) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
(4) Cramer, C. J.; Truhlar, D. G. In *Reviews in Computational Chemistry*, Vol. 6, Boyd, D. B.; Lipkowitz, K. B., Eds.; VCH: New York, 1995; p. 1.
(5) Rivail, J.-L.; Rinaldi, D. In *Computational Chemistry: Reviews of Current Trends*, Vol. 1; Leszczynski, J., Ed.; World Scientific: Singapore, 1996; p. 139.
(6) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305, 9901(E).

(7) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213.

(8) Cramer, C. J.; Truhlar, D. G. *J. Comp. Chem.* **1992**, *13*, 1089.

(9) Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 629.

(10) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comp. Chem.* **1995**, *16*, 422.

(11) Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 1057.

(12) Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1995**, *99*, 7137.

(13) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122.

(14) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385.

(15) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.

(16) Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1996**, *61*, 8720.

(17) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1997**, *101*, 2061.

(18) Giesen, D. J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1997**, *98*, 85.

(19) (a) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 7147. (b) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. to be published.

(20) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B*, to be published.

(21) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. to be published.

(22) (a) Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* to be published. (b) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. unpublished.

(23) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 371.

(24) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.

(25) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, D.C., 1982.

(26) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* **1986**, *7*, 230.

(27) Brooks, C. R.; Bruccoleri, R. E.; Olafson, B. D.; States, B. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187.

(28) Hagler, A. T.; Ewig, C. S. *Comput. Phys. Commun.* **1994**, *84*, 131.

(29) Allinger, N. L.; Zhou, X. F.; Bergsma, J. *J. Mol. Struct. (Theochem)* **1994**, *118*, 69.

(30) (a) Berendsen, H. J. C. In *Computer Simulation of Biomolecular Systems*, Vol. 2; van Gunsteren, W. F., Weiner, P. K., Wilkonson, A. J., Eds.; ESCOM: Leiden, 1993; p. 161. (b) Smith, P. E.; van Gunsteren, W. F. *ibid.*, p. 182.

(31) Smith, P. E.; Pettit, B. M. *J. Chem. Phys.* **1991**, *95*, 8430.

(32) (a) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100. (b) Tasaki, K.; McDonald, S.; Brady, J. W. *J. Comp. Chem.* **1993**, *14*, 278. (c) Guernot, J.; Kollman, P. A. *J. Comp. Chem.* **1993**, *14*, 295.

(33) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *351*.

(34) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486.

(35) Harrison, S. W.; Nolte, N. J.; Beveridge, D. *J. Phys. Chem.* **1976**, *80*, 2580.

(36) (a) Rivail, J. L.; Terryn, B. *J. Chim. Phys.* **1982**, *79*, 1. (b) Rinaldi, D. *Comput. and Chem.* **1982**, *6*, 155.

(37) Hoijtink, G. J.; de Boer, E.; Van der Meij, P. H.; Weijland, W. P. *Rec. Trav. Chim. Pays-Bas* **1956**, *75*, 487.

(38) Peradejordi, F. *Cah. Phys.* **1963**, *17*, 343.

(39) Jano, I. *Compt. Rendu Acad. Sci. Paris* **1965**, *261*, 103.

(40) Kozaki, T.; Morihasi, M.; Kikuchi, O. *J. Am. Chem. Soc.* **1989**, *111*, 1547.

(41) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164.

(42) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.

(43) Bucher, M.; Porter, T. L. *J. Phys. Chem.* **1986**, *90*, 3406.

(44) Schaefer, M.; Froemmel, C. *J. Mol. Biol.* **1990**, *216*, 1045.

(45) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(46) Dewar, M. J. S.; Zoebisch, E. G. *J. Mol. Struct. (Theochem)* **1989**, *187*, 1.

(47) Dewar, M. J. S.; Jie, C. *J. Mol. Struct. (Theochem)* **1989**, *187*, 1.

(48) Dewar, M. J. S.; Yate-Ching, Y. *Inorg. Chem.* **1990**, *29*, 3118.

(49) Stewart, J. J. P. *J. Comp. Chem.* **1989**, *10*, 221.

(50) Stewart, J. J. P. *Rev. Comp. Chem.* **1990**, *1*, 45.

(51) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.

(52) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.

(53) Simons, J.; Nichols, J. *Quantum Mechanics in Chemistry*; Oxford University Press: New York, 1997.

(54) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.

(55) Feller, D.; Davidson, E. R. *Rev. Comp. Chem.* **1990**, *1*, 1.

(56) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(57) Löwdin, P.-O. *J. Chem. Phys.* **1950**, *18*, 365.

(58) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894.

(59) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 431.

(60) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* in press.

(61) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

(62) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81*, 391, **1996**, *93*, 315(E).

(63) Thiel, W.; Voityuk, A. A. *Int. J. Quantum Chem.* **1993**, *44*, 807.

(64) (a) Thiel, W.; Voityuk, A. A. *J. Mol. Struct. (Theochem)* **1994**, *313*, 141. (b) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616. (c) Thiel, W. *Adv. Chem. Phys.* **1996**, *93*, 703.

(65) Hine, J.; Mookerjee, P. K. *J. Org. Chem.* **1975**, *40*, 287.

(66) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Sol. Chem.* **1981**, *10*, 563.

(67) Pearson, R. G. *J. Am. Chem. Soc.* **1986**, *108*, 6109.

(68) Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. *J. Chem. Soc. Perkin Trans. II* **1990**, *291*.

(69) Leo, A. J. *MedChem Database*; BioByte Corp.: Claremont, CA, 1994.

(70) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.

(71) Abraham, M. H. *Chem. Soc. Rev.* **1993**, *22*, 73.

(72) (a) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281. (b) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* in press.

(73) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560.

(74) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T; Still, W. C. *J. Comp. Chem.* **1990**, *11*, 440. We used MACROMODEL–version 6.0. http://www.columbia.edu/cu/chemistry/mmod/mmod.html

(75) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

(76) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1990**, *112*, 8293.

(77) Allinger, N. L.; Yan, L. *J. Am. Chem. Soc.* **1993**, *115*, 11918.

(78) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W., Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. AMSOL–version 6.1.1, Oxford Molecular Group: Campbell, CA, 1997.
    http://comp.chem.umn.edu/amsol
    http://www.oxmol.co.uk/prods/amsol/

(79) AMSOL–version 6.5

(80) G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, OMNISOL–version 1.01, University of Minnesota, Minneapolis, MN 55455
    http://comp.chem.umn.edu/omnisol

(81) AMPAC, Semichem. Inc. Shawnee, KS.
    http://www.semichem.com/ampac.html
    http://comp.chem.umn.edu/ampacm

(82) DGAUSS, a module of UNICHEM, Oxford Molecular Group, Oxford, UK.
    http://www.oxmol.com/prods/unichem/cap/dgauss.html
    http://comp.chem.umn.edu/dgsol

(83) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Gill, P. M. W., Johnson, B. G., Robb, M. A., Cheeseman, J. R., Keith, T. A., Petersson, G. A., Montgomery, J. A., Raghavachari, K., Al-Laham, M. A., Zakrzewski, V. G., Ortiz, J. V., Foresman, J. B., Peng, , C. Y., Ayala, P. A., Wong, M. W., Andres, J. L., Replogle, E. S., Gomperts, R., Martin, R. L., Fox, D. J., Binkley, J. S., Defrees, D. J., Baker, J., Stewart, J. P., Head-Gordon, M., Gonzalez, C., and Pople, J. A., GAUSSIAN94, Gaussian, Inc., Pittsburgh, PA, 1995.
    http://www.gaussian.com/
    http://comp.chem.umn.edu/gsm

(84) 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. *J. Comput. Chem.* **1993**, *14*, 1347.
    http://www.msg.ameslab.gov/GAMESS/GAMESS.html

(85) Li, J.; Cramer, C. J.; Truhlar, D. G. gamesol–version 1.0, University of Minnesota, Minneapolis, MN, 1998.
    http://comp.chem.umn.edu/gamesol