

# GB/SA water model for the Merck molecular force field (MMFF)

Ailan Cheng,\* Scott A. Best,† Kenneth M. Merz Jr.,\*‡ and Charles H. Reynolds†

\*Pharmacoepia, Inc., Princeton, New Jersey, USA

†Rohm and Haas Company, Spring House, Pennsylvania, USA

‡Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania, USA

*A revised generalized Born/surface area (GB/SA) continuum solvation model has been developed for water that is compatible with the Merck molecular force field (MMFF). This model gives free energies of aqueous solvation that are comparable in accuracy to the original water model when the OPLS\* force field is employed. The average unsigned error in aqueous  $\Delta G_{\text{sol}}$  using the new water model and MMFF is 0.62 kcal/mol for a training set of 82 solutes compared to 1.24 kcal/mol for the original GB/SA water model and MMFF. The average unsigned errors for 47 neutral solutes outside the training set and 10 ions are 0.96 and 2.32 kcal/mol, respectively. By comparison, the average errors for the test set and ions using the original GB/SA water model are 1.76 and 5.32 kcal/mol. This revised parameter set provides a more accurate representation of aqueous solvation for use with MMFF. © 2000 by Elsevier Science Inc.*

**Keywords:** continuum solvation model, generalized Born/surface area, Merck molecular force field, water solvation free energy

## INTRODUCTION

Solvation plays an important role in many aspects of chemistry. Solute properties that may be influenced by the solvent medium include electronic configuration, conformation, reactivity, partitioning between solvent phases, and free energies of association.<sup>1</sup> Numerous approaches have been taken to include solvation effects in molecular simulations. Explicit inclusion of solvent in molecular dynamics (MD) and Monte Carlo (MC) simulations have probably been the most widely used methods for studying solvent properties and their effects.<sup>2–4</sup> However, using explicit solvent molecules in these simulations dramatically increases the size and computational expense of the calculation and greatly complicates the task of assuring suffi-

cient sampling of the potential energy surface. Although these all-atom calculations are general and can reproduce many solution state properties, the associated computational expense has led to the development of more efficient approaches, in particular continuum solvation models.<sup>5–22</sup>

The most rigorous continuum methods employ numerical solutions to the Poisson-Boltzmann (PB) equation.<sup>19,23–30</sup> These models only compute one, albeit typically the largest, component of the solvation free energy, the electrostatic solvent polarization energy. The PB equation can be adapted into a form that allows calculation of molecular solvation free energies, as evident in the work of Honig and others.<sup>20,29–34</sup> Recent efforts have focused on developing inexpensive, approximate continuum models that do not sacrifice the accuracy of the PB equation. These approximate continuum models have an advantage over simulations that employ explicit solvent molecules in that they are much less computationally intensive yet often yield accurate results. In continuum methods, the solvent is no longer treated explicitly, but as an ensemble averaged continuous medium having properties of the real solvent.<sup>5,6</sup> Continuum solvation models have been developed for use with both quantum mechanical and empirical force field methods. The semiempirical SMX continuum models developed by Truhlar and Cramer and others<sup>13,35–41</sup> and the generalized Born/surface area (GB/SA) model of Still and coworkers<sup>12,42,43</sup> are both examples of such models.

Perhaps the most important solvent to address using continuum methods is water. It has certainly received the most attention.<sup>5,6</sup> The first GB/SA model<sup>12</sup> was developed to model water. This model gives excellent results when used with either OPLS\* or AMBER\*, but gives much poorer results in conjunction with some other force fields, such as the Merck molecular force field (MMFF).<sup>44–48</sup> This can be illustrated with an example from Table 4. Acetone has an experimental free energy of aqueous solvation of  $-3.9$  kcal/mol. An OPLS\* calculation using GB/SA reproduces this result very well with a calculated  $\Delta G_{\text{aq}} = -3.2$  kcal/mol. By comparison, a similar calculation using MMFF and the Still et al. GB/SA water model gives a predicted  $\Delta G_{\text{aq}}$  that is significantly in error at  $-7.5$  kcal/mol. Similar results are found for other ketones,

Corresponding author: C.H. Reynolds, R.W. Johnson Pharmaceutical Research Institute, Welsh and McKean Roads, Spring House, PA 19477-0776, USA.

E-mail address: CREynol1@prius.jnj.com

esters, amides and amines where polar groups are present. This is not surprising given the difference in the charges used by OPLS\*, AMBER\*, and MMFF. In fact, this strong dependence on the quality of the GB/SA solvation free energies with respect to the force field charges is well documented.<sup>49</sup>

We were interested in developing a GB/SA-MMFF water model for several reasons. First, MMFF has been parameterized for a wide variety of functional groups, such as those commonly found in pharmaceuticals and agrochemicals. It also has been parameterized to model both simple organic compounds and macromolecular systems such as proteins. Finally, MMFF utilizes an empirical scheme for calculating charges that makes atomic charges generally available and mimic electrostatic potential (ESP) charges calculated quantum mechanically utilizing the HF/6-31G\* basis set.<sup>50,51</sup> This provides additional flexibility for modeling compounds with unusual functionality, because 6-31G\*(ESP) partial charges may be generated for atoms in compounds where the MMFF charges are deficient.

## COMPUTATIONAL PROCEDURE

The data set consisted of 129 small organic molecules covering a wide variety of functionalities. The data were randomly separated into training set and test set. The experimental aqueous solvation free energies were taken from the work of Cramer and Truhlar.<sup>52</sup> All aqueous free energies of solution  $\Delta G_{aq}$  were computed using the GB/SA solvation model and simple geometry minimization with the MMFF. This is justified, because the molecules in the data set are rather small and rigid, and the lowest energy conformation is generally the same in the gas and solvated phases. Ensemble averaging of multiple low-energy configurations might be necessary when dealing with much larger and more flexible molecules, where many low-energy conformations exist. All calculations were carried out using GB/SA and MMFF as implemented in MacroModel version 6.0 or 6.5.<sup>53</sup> The atomic surface tension terms ( $\sigma$ ) were fit using multiple linear regression in the JMP statistical analysis package.<sup>54</sup> The  $\sigma$  values were fit with a forced intercept of 0.

## GB/SA METHOD

In the GB/SA model,<sup>12,42</sup> the total solvation free energy ( $\Delta G_{sol}$ ) can be expressed as the sum of a solvent cavity term ( $\Delta G_{cav}$ ), a solute-solvent van der Waals term ( $\Delta G_{vdw}$ ), and a solute-solvent electrostatic term ( $\Delta G_{pol}$ ):

$$G_{sol} = G_{cav} + G_{vdw} + G_{pol} \quad (1)$$

The first two terms account for the free energy for forming a cavity in the solvent and for any dispersion interactions, while the third term takes into consideration the electrostatic free energy contribution. There are two equations that are used to calculate the total solvation free energy for a solute. The first equation takes into consideration  $G_{vdw}$  and  $G_{cav}$ . It has been shown that  $G_{vdw}$  and  $G_{cav}$  are directly proportional to the solvent-accessible surface area (SA) of the solute and can be computed as the sum of the atomic surface area contributions.<sup>8,12,55–58</sup> This is given in Equation 2

$$G_{cav} + G_{vdw} = \sum \sigma_k SA_k \quad (2)$$

where  $SA_k$  is the total solvent-accessible surface area of atom  $k$ , which is a function of the solvent radius, and  $\sigma_k$  is an empirically derived surface tension term dependent on the atom type.

The second equation takes into account electrostatic interactions and is modeled after the generalized Born equation and modified to approximate the PB equation (since the ionic

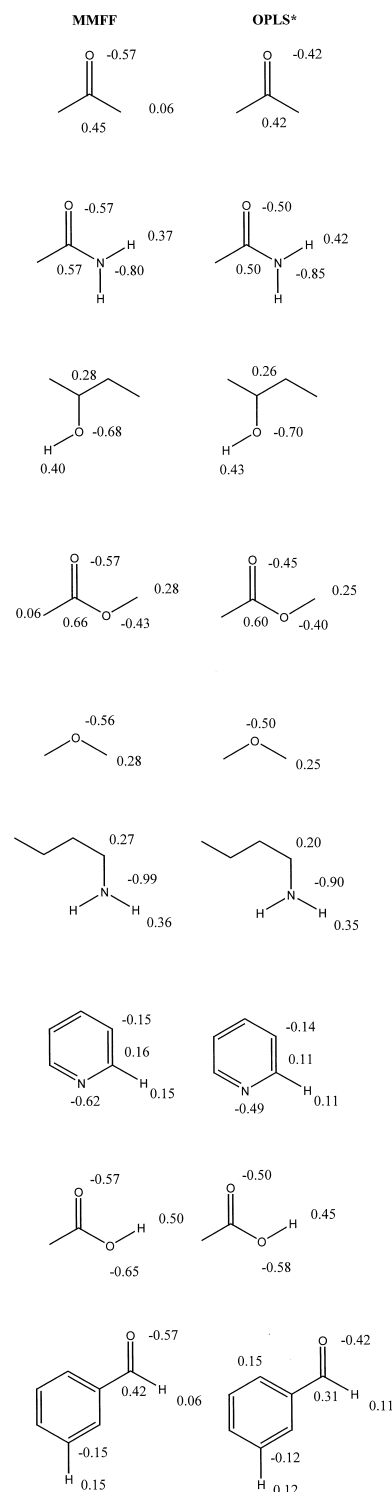


Figure 1. MMFF and OPLS\* partial charges for some representative compounds.

**Table 1. Modified van der Waals (vdw) radii**

Atom type	vdw Radius (Å)	Original value (Å)
O sp <sup>2</sup> (aldehyde)	1.700	1.480
O sp <sup>2</sup> (amide)	1.480	1.480
O sp <sup>2</sup> (other)	1.662	1.480
N sp <sup>3</sup>	1.625 <sup>a</sup>	1.7063
N sp <sup>2</sup> (amide)	1.625 <sup>b</sup>	1.7063
N sp <sup>2</sup> (other)	1.7063	1.7063
N sp	1.900	1.600
F	1.740	1.3665
H polar	1.150	1.2075

<sup>a</sup> OPLS\* value.<sup>b</sup> Using N sp<sup>3</sup> radius for the amide N.**Table 2. Values of P<sub>1</sub>–P<sub>5</sub> for the GB/SA-MMFF and the original Still water continuum model**

Scaling Parameter	GB/SA-MMFF	Original value
P <sub>1</sub>	0.123	0.073
P <sub>2</sub>	0.921	0.921
P <sub>3</sub>	6.211	6.211
P <sub>4</sub>	15.236	15.236
P <sub>5</sub>	1.254	1.254

**Table 3. Optimized σ values**

Atom type	σ (kJ/mol)	Original value (kJ/mol)
CH <sub>3</sub>	0.04544	0.04000
CH <sub>2</sub>	0.02259	0.03000
C <sup>a</sup>	0.02989	0.03000
F	0.12398	0.08000
Cl	0.02808	0.02000
I	−0.01731	−0.03000

<sup>a</sup> All other carbon atom types.

strength is assumed to be zero, this defaults to Poisson equation) as shown in Equation 3:

$$G_{pol} = 166.0 \left( 1 - \frac{1}{\epsilon} \right) \sum_{i=1}^n \sum_{j=1}^n \frac{q_i q_j}{(r_{ij}^2 + \alpha_{ij}^2 e^{-D_{ij}})^{0.5}} \quad (3)$$

with  $\alpha_{ij} = \sqrt{\alpha_i \alpha_j}$  and  $D_{ij} = r_{ij}^2 / (2\alpha_{ij})^2$ ,  $r_{ij}$  being the distance between atoms  $i$  and  $j$  having charges  $q_i$  and  $q_j$ ,  $\epsilon$  being the dielectric constant, and  $\alpha_i$  being the effective Born radius. Still and coworkers<sup>42</sup> have adopted an analytical solution to this equation, which is given in Equation 4

$$G_{pol,i} = \frac{-166.0}{R_{vdW-i} + \varphi + P_1} + \sum^{stretch} \frac{P_2 V_j}{r_{ij}^4} + \sum^{bend} \frac{P_3 V_j}{r_{ij}^4} + \sum^{nonbonded} \frac{P_4 V_j CCF}{r_{ij}^4} \quad (4)$$

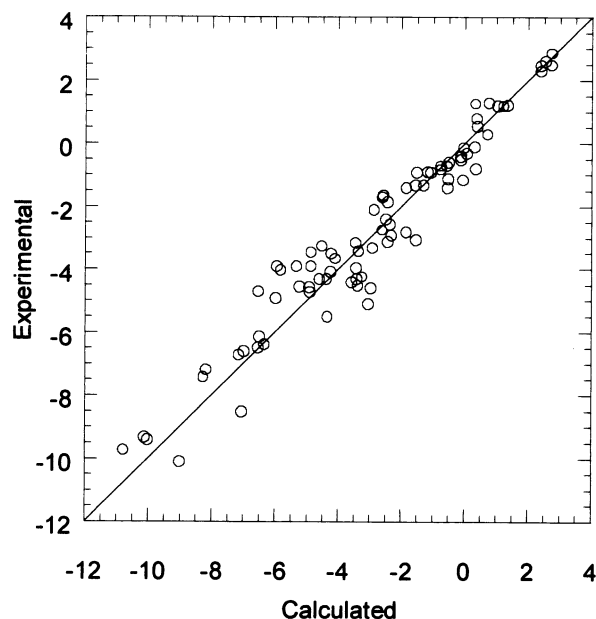


Figure 2. Plot of experimental vs computed  $\Delta G_{sol}$  values for the training set of 82 compounds.

where  $G_{pol,i}$  is the polarization energy of atom  $i$ ,  $r_{ij}$  is the distance between atoms  $i$  and  $j$  (Å),  $R_{vdW-i}$  is the van der Waals radius of atom  $i$  (Å),  $\varphi$  is the dielectric offset,  $V_j$  is the volume of atom  $j$  (Å<sup>3</sup>),  $P_1$  is a single atom scaling factor,  $P_2$  is the stretch scaling factor,  $P_3$  is the bend scaling factor,  $P_4$  is the nonbonded scaling factor,  $P_5$  is the soft cutoff parameter, and CCF is the close contact function for nonbonded interactions (see the original work for the expression for CCF<sup>42</sup>).

## OPTIMIZATION OF THE MODEL

The parameterization of the GB/SA water continuum model using the MMFF was carried out using a protocol analogous to the one used to parameterize the GB/SA octanol model<sup>59</sup> with a few significant exceptions. Unlike the octanol parameterization, the total solvation free energy was optimized rather than initially optimizing the polarization free energy and then optimizing the cavitation and van der Waals terms. Because the primary difference between OPLS\* and MMFF, relative to the GB/SA model, is the origin of the atomic charges, we only optimized one electrostatic parameter ( $P_1$ ).  $P_1$  has the most direct effect on the effective Born radius and, therefore, electrostatic contribution for a given atomic partial charge. Parameters  $P_2$  to  $P_5$  were assigned to the values determined originally by Still and coworkers.<sup>42</sup> Because they were parameterized to fit the PB equation, they should be reasonably transferable.

In the development of our water model for MMFF, we used the original  $P_1$  value as a starting point for a line search of  $P_1$ , initially using the surface tension terms ( $\sigma$ ) from the original water model. At each point in the optimization of  $P_1$ , the nonpolar atomic surface tension terms also were optimized using linear regression. Optimization of the atomic surface tension terms ( $\sigma_k$ ) was carried out using Equation 5, which is derived from Equation 1.

**Table 4. Experimental and MMFF-GB/SA aqueous free energies of solvation (kcal/mol) for compounds in the training set**

Solute	$\Delta G_{\text{aq}}(\text{exp})$	$\Delta G_{\text{aq}}(\text{original})^a$	Error	$\Delta G_{\text{aq}}(\text{revised})^b$	Error
2-Butanol	-4.51	-3.94	0.57	-3.36	1.15
n-Butanol	-4.72	-5.25	-0.53	-4.88	-0.16
n-Octanol	-4.09	-4.39	-0.30	-4.23	-0.14
t-Butanol	-4.58	-3.66	0.92	-2.94	1.64
Cyclohexane	1.23	1.79	0.56	1.35	0.12
n-heptane	2.62	2.65	0.03	2.57	-0.05
Isobutane	2.30	2.14	-0.16	2.42	0.12
n-Hexane	2.48	2.44	-0.04	2.41	-0.07
Neopentane	2.50	2.41	-0.09	2.74	0.24
n-Octane	2.89	2.85	-0.04	2.73	-0.16
n-butylamine	-4.30	-2.45	1.85	-3.38	0.92
Dimethylamine	-4.30	-3.81	0.49	-4.36	-0.06
Methylamine	-4.56	-4.12	0.44	-4.90	-0.34
Piperidine	-5.10	-2.01	3.09	-3.02	2.08
n-Propylamine	-4.40	-2.67	1.73	-3.55	0.85
Triethylamine	-3.04	-1.24	1.80	-1.53	1.51
Trimethylamine	-3.23	-4.30	-1.07	-4.50	-1.27
Diethylether	-1.63	-3.19	-1.56	-2.56	-0.93
THF	-3.50	-4.42	-0.92	-4.21	-0.71
THP	-3.12	-3.56	-0.44	-3.43	-0.31
2-Butanone	-3.64	-6.41	-2.77	-4.09	-0.45
3-Pentanone	-3.41	-5.38	-1.97	-3.35	0.06
Acetone	-3.90	-7.47	-3.57	-4.86	-0.96
Cyclopentanone	-4.54	-7.28	-2.74	-5.23	-0.69
Acetic acid	-6.70	-9.21	-2.51	-7.13	-0.43
Butanoic acid	-6.36	-8.13	-1.77	-6.33	0.03
Propanoic acid	-6.47	-8.32	-1.85	-6.52	-0.05
Acetamide	-9.71	-11.39	-1.68	-10.76	-1.05
N-methylacetamide	-10.10	9.74	0.36	-8.99	1.11
N,N-dimethylacetamide	-8.50	-7.91	0.59	-7.02	1.48
N-Propylamide	-9.40	-10.51	-1.11	-10.00	-0.60
4-Methylpyridine	-4.90	-6.96	-2.06	-5.97	-1.07
Anthracene	-4.23	-4.01	0.22	-3.23	1.00
Benzene	-0.90	-1.96	-1.06	-1.49	-0.59
Ethylbenzene	-0.80	-1.24	-0.44	-0.77	0.03
Napthalene	-2.39	-3.11	-0.72	-2.48	-0.09
Phenol	-6.60	-7.63	-1.03	-6.97	-0.37
Phenanthrene	-3.95	-4.18	-0.23	-3.40	0.55
Pyridine	-4.70	-7.43	-2.72	-6.51	-1.81
Toluene	-0.90	-1.62	-0.72	-1.06	-0.16
Ethyl acetate	-3.10	-4.38	-1.28	-2.41	0.69
Ethyl propanate	-2.80	-3.53	-0.73	-1.83	0.97
Methyl acetate	-3.30	-4.94	-1.64	-2.90	0.40
Methyl propanate	-2.90	-4.07	-1.17	-2.30	0.60
1,1-Difluoroethane	-0.10	-3.05	-2.95	0.32	0.42
1,2-Dibromoethane	-2.10	-4.01	-1.91	-2.86	-0.76
1,2-Dichloroethane	-1.70	-4.61	-2.91	-2.59	-0.89
1-Bromobutane	-0.41	-0.69	-0.28	-0.10	0.31
1-Chlorobutane	-0.13	-1.06	-0.93	-0.01	0.12
2-Iodopropane	-0.50	-0.59	-0.09	-0.11	0.39
Bromoethane	-0.70	-1.21	-0.51	-0.54	0.16
Chlorethane	-0.60	-1.63	-1.03	-0.47	0.11
Fluoroform	0.80	-3.95	-4.75	0.38	-0.42

(Continued)

Table 4. (Continued)

Solute	$\Delta G_{\text{aq}}(\text{exp})$	$\Delta G_{\text{aq}}(\text{original})^a$	Error	$\Delta G_{\text{aq}}(\text{revised})^b$	Error
Iodoethane	-0.70	-1.12	-0.42	-0.75	-0.05
Methyl iodide	-0.89	-1.55	-0.66	-1.17	-0.28
Trifluoroethanol	-4.30	-8.78	-4.48	-4.58	-0.28
Benzenethiol	-2.55	-2.14	0.41	-2.34	0.21
Diethyl sulfide	-1.30	-1.26	0.04	-1.28	0.02
Methyl ethyl sulfide	-1.40	-1.79	-0.39	-1.83	-0.43
1-Hexyne	0.30	0.64	0.34	0.72	0.42
2-Methylpropene	1.20	0.70	-0.50	1.06	-0.14
Acetonitrile	-3.90	-8.77	-4.87	-5.93	-2.03
Ethylene	1.27	0.19	-1.08	0.34	-0.93
Propanenitrile	-3.90	-7.99	-4.09	-5.31	-1.41
Propyne	-0.30	-0.21	0.09	0.08	0.38
Propene	1.30	0.51	-0.79	0.77	-0.53
1,2-Ethanediol	-9.30	-10.77	-1.47	-10.12	-0.82
4-Methylphenol	-6.14	-7.23	-1.09	-6.48	-0.34
Fluorobenzene	-0.78	-1.34	-0.56	0.36	1.14
Aniline	-5.49	-4.98	0.51	-4.33	1.16
Benzaldehyde	-4.02	-9.33	-5.31	-5.81	-1.79
Chlorobenzene	-1.12	-1.51	-0.39	-0.51	0.61
Cyclopentane	1.20	1.64	0.44	1.24	0.04
Cyclopentene	0.56	0.50	-0.06	0.40	-0.16
Di-methyl-disulfide	-1.83	-1.89	-0.06	-2.40	-0.57
Ethanethiol	-1.30	-1.13	0.17	-1.55	-0.25
Morpholine	-7.17	-7.88	-0.71	-8.16	-0.99
o-Chlorotoluene	-1.15	-1.00	0.15	-0.03	1.12
p-Bromotoluene	-1.39	-1.28	0.11	-0.52	0.87
Piperazine	-7.40	-6.64	0.76	-8.26	-0.86
Propanal	-3.44	-8.12	-4.68	-4.86	-1.42
Thioanisole	-2.73	-2.73	0.00	-2.61	0.12
RMS error			1.78		0.81
Average			-0.85		-0.06
Unsigned Average			1.24		0.62
Largest error			-5.31		2.08

<sup>a</sup>Original GB/SA water model and MMFF.<sup>b</sup>GB/SA water model that has been revised for MMFF.

$$\Delta G_{\text{sol}} - \Delta G_{\text{pol}} = \Delta G_{\text{cav}} + \Delta G_{\text{vdw}} = \sum \sigma_k S A_k \quad (5)$$

Simply optimizing  $P_1$  was not sufficient to produce an accurate water model for MMFF. This is due to the fact that the MMFF and OPLS\* charges are not linearly related, but vary with respect to certain polar groups. The MMFF and OPLS\* partial charges are compared for a few representative compounds in Figure 1. For example, it is clear that the difference between MMFF and OPLS\* partial charges on a carbonyl oxygen depends on whether the carbonyl oxygen is part of a ketone or amide functionality. This degree of variation can have a dramatic effect on the solvation free energies.<sup>49</sup>

In order to deal with problems such as the carbonyl oxygen, we also scaled some of the van der Waals radii in our revised water model in order to account for these differences in partial charge. This is purely an empirical expedient in order to bring the total  $G_{\text{pol}}$  term using the MMFF charges into line with the values necessary to reproduce experimental free energies of aqueous solvation. We elected to make the fewest number of

changes to the van der Waals radii necessary to obtain an acceptable model. One might alternatively decide to use all of the polar radii as adjustable parameters. Dominy and Brooks<sup>60</sup> have fit a GB/SA water model for the CHARMM force field using a scaling parameter for the van der Waals radii. In the final model we adopted modified van der Waals radii relative to the radii in version 6.5 of MacroModel for seven atom types (Table 1). In some cases, we simply reset radii that Still and coworkers had scaled back to their original OPLS\* values. For example, we reset the N3 atom type radius back to the OPLS\* value of 1.625 Å, and the radius of polar hydrogens, H2 and H3, were set to the value used in the original GB/SA paper of 1.15 Å. We gave the N2 atom type for the nitrogen in amides the same radius as N3. The radii for carbonyl oxygens were scaled up in size except for the carbonyl oxygen in amides. Finally, the radii for N1 and F were increased in order to compensate for their MMFF charges and reduce their contributions to the total calculated  $G_{\text{pol}}$ .



**Table 5. Experimental and MMFF-GB/SA aqueous free energies of solvation (kcal/mol) for compounds outside the training set**

Solute	$\Delta G_{\text{aq}}(\text{exp})$	$\Delta G_{\text{aq}}(\text{original})^a$	Error	$\Delta G_{\text{aq}}(\text{revised})^b$	Error
1-Pentyne	0.00	0.42	0.42	0.60	0.56
1,2-Dimethoxyethane	-4.80	-7.12	-2.32	-6.01	-1.21
2-Bromopropane	-0.50	-0.65	-0.15	-0.27	0.64
2-Methoxyethanol	-6.80	-8.98	-2.18	-8.11	-1.30
3-Pentanol	-4.30	-3.11	1.19	-2.62	1.67
4-Heptanone	-2.90	-5.01	-2.11	-2.97	-0.07
Butanal	-3.20	-7.50	-4.30	-5.18	-1.28
Dipropylamine	-3.70	-1.12	2.58	-1.83	1.88
Methylbutanoate	-2.80	-3.87	-1.07	-2.10	0.69
o-Xylene	-0.90	-1.40	-0.50	-0.69	0.13
p-Xylene	-0.80	-1.28	-0.48	-0.55	0.17
Pentanoic acid	-7.00	-7.88	-0.88	-6.14	0.86
Propane	1.96	1.82	-0.14	1.96	-0.01
n-Butane	2.08	2.03	-0.05	2.11	0.03
n-Pentane	2.33	2.24	-0.09	2.26	-0.07
2-Methylpentane	2.52	2.54	0.02	2.73	0.19
2,4-Dimethylpentane	2.88	2.79	-0.09	3.11	0.21
Cyclopropane	0.75	1.37	0.62	1.00	0.28
1-Butene	1.38	0.86	-0.52	1.08	-0.35
1-Pentene	1.66	0.18	-1.48	1.24	-0.47
1-Butyne	-0.16	1.08	1.24	0.42	0.53
Butenyne	0.04	-0.51	-0.55	-0.17	-0.31
Dimethyl ether	-1.92	-4.96	-3.04	-4.12	-2.22
Methyl propyl ether	-1.66	-3.35	-1.69	-2.72	-1.07
1,4-Dioxane	-5.05	-7.63	-2.58	-6.97	-1.90
Ethanal	-3.50	-7.64	-4.14	-6.12	-1.86
Pentanal	-3.03	-8.76	-5.73	-5.26	-1.46
2-Hexanone	-3.29	-5.98	-2.69	-3.71	-0.42
Ethylamine	-4.50	-3.00	1.50	-3.83	0.67
Diethylamine	-4.07	-1.66	2.41	-2.26	1.80
Azetidine	-5.56	-2.46	3.10	-3.31	2.28
Pyrrolidine	-5.48	-2.85	2.63	-3.92	1.58
2-Methylpyridine	-4.77	-6.66	-1.89	-5.62	-0.93
Hydrogen sulfide	-0.70	-1.15	-0.45	-1.32	-1.58
Methanethiol	-1.24	-1.59	-0.35	-1.74	-0.73
Ammonia	-4.29	-5.59	-1.30	-6.22	-1.93
Hydrazine	-9.30	-6.75	2.55	-8.50	0.80
2-Chloropropane	-0.25	-1.02	-0.77	0.27	0.46
Trichloromethane	-1.07	-8.88	-7.81	-5.48	-4.82
Tetrafluoromethane	3.16	-0.54	-3.70	4.96	0.18
Bromobenze	-1.46	-1.71	-0.25	-1.27	0.42
Chlorofluoromethane	-0.77	-4.88	-4.11	-1.54	-1.41
Chloroethene	-0.59	-0.86	-0.27	0.29	0.51
Fluormethane	-0.22	-2.16	-1.94	0.71	0.41
Dibromomethane	-2.11	-4.92	-2.81	-4.59	-1.55
p-Dichlorobenze	-1.01	-1.59	-0.58	0.52	0.92
Hexanoic acid	-6.21	-7.66	-1.45	-5.98	0.23
RMS			2.39		1.29
Average			-0.98		-0.19
Unsigned average			1.76		0.96
Largest error			-7.81		-4.82

<sup>a</sup>Original GB/SA water model and MMFF.<sup>b</sup>GB/SA water model that has been revised for MMFF.

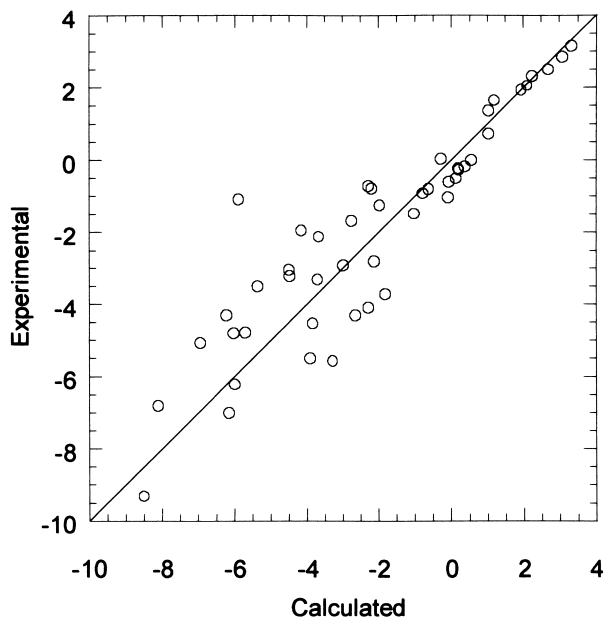


Figure 3. Plot of experimental vs computed  $\Delta G_{\text{sol}}$  values for 47 compounds outside the training set.

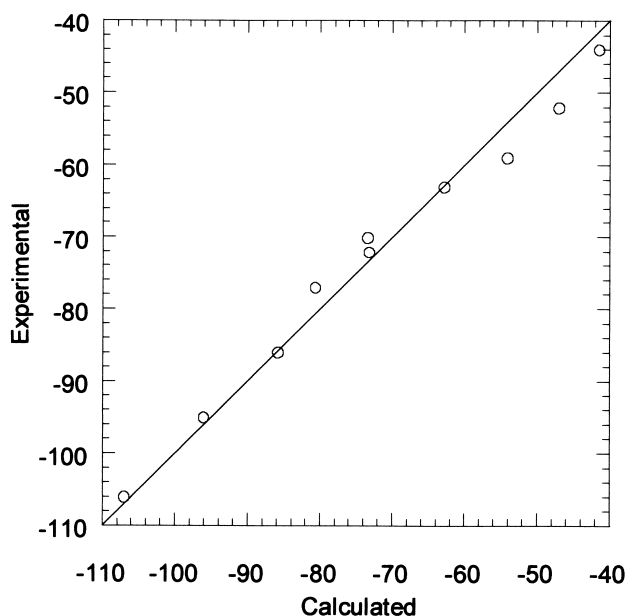


Figure 4. Experimental vs computed  $\Delta G_{\text{sol}}$  values for 10 ions listed in Table 6.

Once the van der Waals radii were set, a final optimization of  $P_1$  was carried out followed by determination of the atomic SASA contributions (Table 2). In the final optimization, the  $\sigma$  values were optimized in two steps. First, the  $\text{CH}_3$ ,  $\text{CH}_2$  and  $\text{C}_{\text{other}}$  values were optimized using the new  $P_1$  and van der Waals radii for a subset of the data that only included hydrocarbons. This was done to ensure that the  $\sigma$  increments for these alkyl groups were not used to absorb other errors in the full training set. Once these  $\sigma$  values were obtained, the  $\text{CH}_3$  and  $\text{CH}_2$   $\sigma$  values were frozen, and  $\sigma$  value of  $\text{C}_{\text{other}}$ , F, Cl, and

I were optimized using the full training set via linear regression. The final  $\sigma$  values are given in Table 3. Addition of  $\sigma$  terms for the polar atoms such as oxygen and nitrogen only had a small effect on the total root mean square (RMS) error and, therefore, were not included in the model. This is consistent with the fact that electrostatics dominate the solvation free energy for the polar atoms and follows the precedent of the original GB/SA water model.

## MODEL RESULTS

The experimental and computed values for  $\Delta G_{\text{sol}}$  are listed in Table 4 for each compound in the training set. Figure 2 is a plot of the calculated  $\Delta G_{\text{sol}}$  vs the experimental  $\Delta G_{\text{sol}}$ . This plot shows good agreement for almost every compound in the training set with an  $r^2 = 0.93$  and an averaged unsigned error of 0.62 kcal/mol. The calculated and experimental free energies of aqueous solvation are given in Table 5 for 47 solutes that are outside the training set. The computed  $\Delta G_{\text{sol}}$  values for these compounds also show good agreement with experiment. The average unsigned error for this test set is 0.96 kcal/mol. A plot of calculated vs experimental free energies of solvation for these compounds is given in Figure 3. We did not note any correlation between the error and the size or flexibility of the molecules. Examination of two different conformations of n-hexane (one all-trans, another folded) shows that the straight chain is the lowest energy configuration in both the gas and solvated phase. Even though the energies of the two conformations differ by 1.6 kcal/mol, the solvation free energies only differ by less than 0.1 kcal/mol. This likely due to the fact that the intramolecular interactions are the same in the gas and solvated phase. For 1,2-ethanediol, an intramolecular hydrogen bond is formed for the lowest energy conformation in the gas phase. This conformation is also the lowest energy structure in the solvated phase. The trans-conformation, which does not allow for intramolecular hydrogen bonding, has a higher energy (2.54 kcal/mol higher in water and 2.91 kcal/mol higher in the gas phase) than the lowest energy cis-conformation, yet the solvation free energy is 0.37 kcal/mol lower. In this case, the lowest energy conformation gives a result that is consistent with the result one would calculate using the ensemble average. However, when dealing with much larger and more flexible molecules with multiple low-energy conformations, ensemble averaging may become necessary.

Finally, we also examined a small set of ions (Table 6, Figure 4) using our revised GB/SA model and MMFF. Of course, electrostatics play an even larger role in charged solutes where the free energies of aqueous solvation can be very large. For example, the experimental  $\Delta G_{\text{aq}}$  values for the 10 ions in Table 6 range from  $-44$  to  $-106$  kcal/mol. Given the much larger magnitude of the free energies of solvation for the cations and anions in Table 6, the unsigned average error of 2.32 kcal/mol is excellent. The calculated  $\Delta G_{\text{aq}}$  for the ions are plotted against their experimental values in Figure 4.

As is evident from the smaller RMS and unsigned average errors, the reparameterized GB/SA model described here is a significant improvement when used with MMFF. The average unsigned error for all 129 neutral solutes is 0.74 kcal/mol, or about half the unsigned average error of 1.43 that results from using the original model with MMFF. The largest improvements are found for a few classes of polar solutes. For example, the ketones, acids, and aldehydes are much improved as are

**Table 6. Experimental and MMFF-GB/SA aqueous free energies of solvation (kcal/mol) for ions**

Solute	$\Delta G_{\text{aq}}$ (exp)	$\Delta G_{\text{aq}}$ (original) <sup>a</sup>	Error	$\Delta G_{\text{aq}}$ (revised) <sup>b</sup>	Error
Acetate	-77.00	-89.83	-12.83	-80.63	-3.65
Ammonium	-86.00	-88.25	-2.25	-85.73	0.27
Methylammonium	-70.00	-75.43	-5.43	-73.34	-3.35
Dimethylammonium	-63.00	-64.54	-1.54	-62.75	0.23
Trimethylammonium	-59.00	-48.59	10.41	-54.05	4.93
Tetramethylammonium	-52.00	-55.69	-3.69	-46.92	5.06
Tetraethylammonium	-44.00	-42.52	1.48	-41.36	2.63
Hydroxide	-106.00	-110.31	-4.31	-106.93	-0.93
Methoxide	-95.00	-99.61	-4.61	-96.04	-1.05
NO <sub>3</sub> <sup>-</sup>	-72.00	-78.60	-6.60	-73.11	-1.11
RMS			6.40		2.91
Average			-2.94		0.30
Unsigned average			5.32		2.32
Largest error			-12.83		5.06

<sup>a</sup>Original GB/SA water model and MMFF.<sup>b</sup>GB/SA water model that has been revised for MMFF.

most of the halogens. There are still a few solutes that produce errors in excess of 2 kcal/mol such as piperidine, acetonitrile, dimethyl ether, azetidine, and trichloromethane. Trichloromethane is the only neutral solute with an error in excess of 2.5 kcal/mol. The error for trichloromethane is very large at -4.8 kcal/mol, albeit smaller than the corresponding error for the original Still model of -7.8 kcal/mol.

## CONCLUSION

We have derived parameters for a revised version of GB/SA that can be used to calculate free energies of aqueous solvation with the MMFF. This GB/SA water model should be broadly applicable to any simulation where one wants to use MMFF and include aqueous solvation effects. Further, this revised GB/SA water model can be used in conjunction with our previously described GB/SA octanol model<sup>59,61</sup> to compute free energies of solvation in water and octanol using a consistent force field model (MMFF). Some potential applications for this model include molecular recognition,<sup>62</sup> ligand-receptor binding,<sup>63</sup> and calculation of partition coefficients.<sup>64</sup>

## REFERENCES

- Reichardt, C. *Solvent Effects in Organic Chemistry*. Verlag Chemie: Weinheim, 1979
- Jorgensen, W.L., and Ravimohan, C.J. Monte Carlo simulation of differences in free energies of hydration. *J. Chem. Phys.* 1985, **83**, 3050-3054
- Rosky, P.J., and Karplus, M. Solvation: A molecular dynamics study of a dipeptide in water. *J. Am. Chem. Soc.* 1979, **101**, 1913-1936
- DeBolt, S.E., and Kollman, P.A. Investigation of structure, dynamics, and solvation in 1-octanol and its water-saturated solution: Molecular dynamics and free-energy perturbation studies. *J. Am. Chem. Soc.* 1995, **117**, 5316-5340
- Cramer, C.J., and Truhlar, D.G. Implicit solvation models: Equilibria, structure, spectra and dynamics. *Chem. Rev.* 1999, **99**, 2161-2200
- Cramer, C.J., and Truhlar, D.G. Continuum solvation models: Classical and quantum mechanical implementations. In: *Reviews in Computational Chemistry, Volume 6*, Lipkowitz, K.B., and Boyd, D. (eds.). VCH, New York, 1995, pp. 1-72
- Eisenberg, D., and McLachlan, A.D. Solvation energy in protein folding and binding. *Nature* 1986, **319**, 199-203
- Ooi, T., Oobatake, M., Nemethy, G., and Scheraga, H.A. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci. U. S. A.* 1987, **84**, 3086-3090
- Miertus, S., Scrocco, E., and Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* 1981, **55**, 117-29
- Kang, Y.K., Nemethy, G., and Scheraga, H.A. Free energies of hydration of solute molecules. 1. Improvement of the hydration shell model by exact computations of overlapping volumes. *J. Phys. Chem.* 1987, **91**, 4105-4109
- Warshel, A., and Russell, S.T. Calculations of electrostatic interactions in biological systems and in solutions. *Rev. Biophys.* 1984, **17**, 283
- Still, W.C., Tempczyk, A., Hawley, R., and Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 1990, **112**, 6127-6129
- Cramer, C.J., and Truhlar, D.G. General parameterized SCF model for free energies of solvation in aqueous solution. *J. Am. Chem. Soc.* 1991, **113**, 8305-8311
- Tomasi, J., and Persico, M. Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent. *Chem. Rev.* 1994, **94**, 2027-2094
- Tomasi, J., Bonaccorsi, R., Cammi, R., Olivares del Valle, F.J. Theoretical chemistry in solution: Some results and perspectives of the continuum methods and in



- particular of the polarizable continuum model. *J. Mol. Struct. (THEOCHEM)* 1991, **234**, 401–424
- 16 Tawa, G.J., Martin, R.L., Pratt, L.R., and Russo, T.V. Solvation free energy calculations using a continuum dielectric model for the solvent and gradient-corrected density functional theory for the solute. *J. Phys. Chem.* 1996, **100**, 1515–1523
  - 17 Chen, J.L., Noodleman, L., Case, D.A., and Bashford, D. Incorporating solvation effects into density functional electronic structure. *J. Phys. Chem.* 1994, **98**, 11059–11068
  - 18 Rashin, A.A., Young, L., and Topol, I.A. Quantitative evaluation of hydration thermodynamics with a continuum model. *Biophys. Chem.* 1994, **51**, 359–374
  - 19 Davis, M.E., and McCammon, J.A. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* 1990, **90**, 509–524
  - 20 Gilson, M., and Honig, B. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Prot. Struct. Funct. Genet.* 1988, **4**, 7–18
  - 21 Schmidt, A.B., and Fine, R.M. A CFF91-based continuum solvation model: solvation free energies of small organic molecules and conformations of the alanine dipeptide in solution. *Mol. Simul.* 1994, **13**, 347–365
  - 22 Klamt, A. Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* 1995, **99**, 2224–2235
  - 23 Warwicker, J., and Watson, H.C. Calculation of the electric potential in the active cleft due to  $\alpha$ -helix dipoles. *J. Mol. Biol.* 1982, **157**, 671–679
  - 24 Klapper, I., Hangstrom, R., Fine, R., Sharp, K., and Honig, B. Focusing of electric fields in the active site of copper-zinc superoxide dismutase: effects of ionic strength and amino-acid modification. *Prot. Struct. Funct. Genet.* 1986, **1**, 47–59
  - 25 Sharp, K., Jean-Charles, A., and Honig, B. A local dielectric constant model for solvation free energies which accounts for solute polarizability. *J. Phys. Chem.* 1992, **96**, 3822–3828
  - 26 Rashin, A.A. Electrostatics of ion-ion interactions in solution. *J. Phys. Chem.* 1989, **93**, 4664–4669
  - 27 Juffer, A.H., Botta, E.F.F., van Keulen, B.A.M., van der Plog, A., Berendsen, H.J.C. The electric potential of a macromolecule in a solvent: A fundamental approach. *J. Comp. Phys.* 1991, **97**, 144–171
  - 28 You, T.J., and Harvey, S.C. Finite element approach to the electrostatics of macromolecules with arbitrary geometries. *J. Comput. Chem.* 1993, **14**, 484–501
  - 29 Gilson, M.K., Sharp, K.A., and Honig, B. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comput. Chem.* 1988, **9**, 327–335
  - 30 Zauhar, R.J., and Morgan, R.S. The rigorous computation of the molecular electric potential. *J. Comput. Chem.* 1988, **9**, 171–187
  - 31 Sitkoff, D., Ben-Tal, N., and Honig, B. Calculation of alkane to water solvation free energies using continuum solven models. *J. Phys. Chem.* 1996, **100**, 2744–2752
  - 32 Gilson, M.K., Rashin, A., Fine, R., and Honig, B. On the calculation of electrostatic interactions in proteins. *J. Mol. Biol.* 1985, **88**, 503–516
  - 33 Zauhar, R.J., and Morgan, R.S. A new method for computing the macromolecular electric potential. *J. Mol. Biol.* 1985, **186**, 815–820
  - 34 Luty, B.A., Davis, M.E., and McCammon, J.A. Electrostatic energy calculations by a finite-difference method: Rapid calculation of charge-solvent interaction energies. *J. Comput. Chem.* 1992, **13**, 768–771
  - 35 Giesen, D.J., Storer, J.W., Cramer, C.J., and Truhlar, D.G. General semiempirical quantum mechanical solvation model for nonpolar solvation free energies. n-Hexadecane. *J. Am. Chem. Soc.* 1995, **117**, 1057–1068
  - 36 Giesen, D.G., Cramer, C.J., and Truhlar, D.G. A semiempirical quantum mechanical solvation model for solvation free energies in all alkane solvents. *J. Phys. Chem.* 1995, **99**, 7137–7146
  - 37 Chambers, C.C., Hawkins, G.D., Cramer, C.J., and Truhlar, D.G. Model for aqueous solvation based on class IV atomic charges and first solvation shell effects. *J. Phys. Chem.* 1996, **100**, 16385–16398
  - 38 Cramer, C.J., and Truhlar, D.G. AM1-SM2 and PM3-SM3 parameterized SCF solvation models for free energies in aqueous solution. *J. Comput. Aided Mol. Design* 1992, **6**, 629–666
  - 39 Giesen, D.J., Chambers, C.C., Cramer, C.J., and Truhlar, D.G. Solvation model for chloroform based on class IV atomic charges. *J. Phys. Chem. B* 1997, **101**, 2061–2069
  - 40 Hawkins, G.D., Cramer, C.J., and Truhlar, D.G. Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* 1996, **100**, 19824–19839
  - 41 Giesen, D.J., Gu, M.Z., Cramer, C.J., and Truhlar, D.G. A universal organic solvation model. *J. Org. Chem.* 1996, **61**, 8720–8721
  - 42 Qiu, D., Shenkin, P.S., Hollinger, F.P., and Still, W.C. The GB/SA continuum model for solvation: A fast analytical method for the calculation of approximate born radii. *J. Phys. Chem. A* 1997, **101**, 3005–3014
  - 43 Edinger, S.R., Cortis, C., Shenkin, P.S., and Friesner, R.A. Solvation free energies of peptides: comparison of approximate continuum solvation models with accurate solution of the Poisson-Boltzmann equation. *J. Phys. Chem. B* 1997, **101**, 1190–1197
  - 44 Halgren, T.A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* 1996, **17**, 490–512
  - 45 Halgren, T.A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* 1996, **17**, 520–552
  - 46 Halgren, T.A. Merck molecular force field. III. Molecular geometries and vibrational parameters for intermolecular interactions. *J. Comput. Chem.* 1996, **17**, 553–586
  - 47 Halgren, T.A. Merck molecular force field. IV. Conformational energies and geometries for MMFF94. *J. Comput. Chem.* 1996, **17**, 587–615
  - 48 Halgren, T.A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* 1996, **17**, 616–641
  - 49 Reddy, M.R., Erion, M.D., Agarwal, A., Viswanadhan, V.N., McDonald, D.Q., and Still, W.C. Solvation free energies calculated using the GB/SA model: Sensitivity

- of results on charge sets, protocols, and force fields. *J. Comput. Chem.* 1998, **19**, 769–780
- 50 Singh, U.C., and Kollman, P.A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* 1984, **5**, 129–145
  - 51 Besler, B.H., Merz, K.M. Jr., and Kollman, P.A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* 1990, **11**, 431–439
  - 52 Hawkins, G.D., Liotard, D.A., Cramer, C.J., and Truhlar, D.G. OMNISOL: Fast prediction of free energies of solvation and partition coefficients. *J. Org. Chem.* 1998, **63**, 4305–4313
  - 53 Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., and Still, W.C. MacroModel: An integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* 1990, **11**, 440
  - 54 *JMP Version 3.15*. SAS Institute Inc., Cary, NC, 1989–1995
  - 55 Hermann, R.B. Theory of hydrophobic bonding. II. the Correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem.* 1972, **76**, 2754–2759
  - 56 Amidon, G.L., Yalkowsky, S.H., Anik, S.T., and Valvani, S.C. Solubility of nonelectrolytes in polar solvents. V. Estimation of the solubility of aliphatic monofunctional compounds in water using a molecular surface area approach. *J. Phys. Chem.* 1975, **79**, 2239–2246
  - 57 Floris, F., and Tomasi, J. Evaluation of the dispersion contribution to the solvation energy: A simple computational model in the continuum approximation. *J. Comput. Chem.* 1989, **10**, 616–627
  - 58 Sitkoff, D., Sharp, K.A., and Honig, B. Accurate calculation of hydration free energy using macroscopic solvent models. *J. Phys. Chem.* 1994, **98**, 1978–1988
  - 59 Best, S.A., Merz, K.M. Jr., and Reynolds, C.H. GB/SA-based continuum solvation model for octanol. *J. Phys. Chem. B* 1997, **101**, 10479–10487
  - 60 Dominy, B.N., and Brooks, C.L. III. Development of a generalized Born model parametrization for proteins and nucleic acids. *J. Phys. Chem. B* 1999, **103**, 3765–3773
  - 61 Best, S.A., Merz, K.M. Jr., and Reynolds, C.H. Free energy perturbation study of octanol/water partition coefficients: comparison with continuum GB/SA calculations. *J. Phys. Chem. B* 1999, **103**, 714–726
  - 62 Luo, R., Head, M.S., Given, J.A., and Gilson, M.K. Nucleic acid base-pairing and N-methylacetamide self-association in chloroform: Affinity and conformation. *Biophys. Chem.* 1999, **78**, 183–193
  - 63 Zou, X., Sun, Y., and Kuntz, I.D. Inclusion of solvation in ligand binding free energy calculations using the generalized-Born model. *J. Am. Chem. Soc.* 1999, **121**, 8033–8043
  - 64 Reynolds, C.H., and Best, S.A. A fast molecular simulation to calculate lipophilicity. *Chemtech* 1998, **28**, 28–34