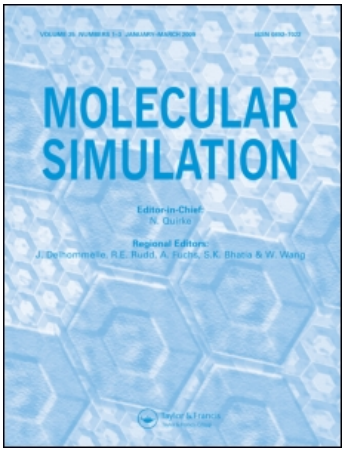


**Molecular Simulation**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

# A CFF 91-based Continuum Solvation Model: Solvation Free Energies of Small Organic Molecules and Conformations of the Alanine Dipeptide in Solution

A. B. Schmidt[a]; R. M. Fine[a]
[a] Biosym Technologies, Inc., Parsippany, NJ

**To cite this Article** Schmidt, A. B. and Fine, R. M.(1994) 'A CFF 91-based Continuum Solvation Model: Solvation Free Energies of Small Organic Molecules and Conformations of the Alanine Dipeptide in Solution', Molecular Simulation, 13: 6, 347 — 365
**To link to this Article: DOI:** 10.1080/08927029408021998
**URL:** http://dx.doi.org/10.1080/08927029408021998

*Molecular Simulation*, 1994, Vol. 13, pp. 347-365
Reprints available directly from the publisher
Photocopying permitted by license only



# A CFF91-BASED CONTINUUM SOLVATION MODEL: SOLVATION FREE ENERGIES OF SMALL ORGANIC MOLECULES AND CONFORMATIONS OF THE ALANINE DIPEPTIDE IN SOLUTION

A.B. SCHMIDT and R.M. FINE†

*Biosym Technologies, Inc. 4 Century Drive, Parsippany, NJ 07054*

*(Received April 1994, accepted May 1994)*

For molecular mechanics simulations of solvated molecules, it is important to use a consistent approach for calculating both the force field energy and the solvation free energy. A continuum solvation model based upon the atomic charges provided with the CFF91 force field is derived. The electrostatic component of the solvation free energy is described by the Poisson-Bolzmann equation while the nonpolar comonent of the solvation energy is assumed to be proportional to the solvent accessible surface area of the solute. Solute atomic radii used to describe the interface between the solute and solvent are fitted to reproduce the energies of small organic molecules. Data for 140 compounds are presented and compared to experiment and to the results from the well-characterized quantum mechanical solvation model AM1-SM2. In particular, accurate results are obtained for amino acid neutral analogues (mean unsigned error of 0.3 kcal/mol). The conformational energetics of the solvated alanine dipeptide is discussed.

KEY WORDS: Molecular mechanics, solvation free energy, solvation continuum model, small organic molecules, alanine dipeptide



## 1 INTRODUCTION

Solvent effects play a crucial role in determining properties of solutes in aqueous solution. Solute properties affected range from their electronic configurations to their conformational energetics and rates of chemical reactions [1-4]. Despite this well-established criticality, there has been relatively little emphasis placed on the development of well-parameterized and well-tested methods for including solvent effects into the general computational framework of molecular mechanical and quantum mechanical studies on molecular solutes until quite recently. This has been due in part to the possibility of simply including a large number of explicit water molecules within the calculation to capture the free-energetic effects of solute-solvent interactions on solute properties.

For molecular dynamics studies of solutes the use of explicit water molecules is straightforward but expensive. Depending on the nature of the solute the use of explicit water can demand that an equal if not larger number of computer cycles be spent on simulating solvent degrees of freedom than on simulating the solute of interest. However, the use of explict water molecules is incompatible with efficient

†Author to whom correspondence should be addressed.

conformational search techniques designed to find low-energy conformations of solutes by generating random or systematic moves in solute internal degrees of freedom. In these methods small changes in solute internal degrees of freedom can result in large changes in atomic cartesian coordinates. These large changes generate energetically unfavorable atomic overlaps with surrounding explicit solvent molecules, resulting in rejection of the move. Quantum mechanical studies of solutes [5–10] can similarly make use of explicit waters in their description of solvent. Recent studies have moved beyond the use of statically placed water molecules to represent the solvent as an external electrostatic field to the use of molecular mechanical methods to provide an appropriate ensemble-averaged description of solute-solvent interactions. While this provides a better match between the number of computer cycles spent on solute description vs. solvent simulation, the incremental cost is still significant while questions remain on validity of coupling of those portions of the system described using molecular mechanics to those portions of the system described using quantum mechanics.

An attractive alternative to explicit water molecules is the description of the solvent as an ensemble-averaged continuum. Such models are not new: they have a history which spans a significant portion of this century. For electrostatic solvation effects this past decade has seen the replacement of simple spherical models [11–13] with the cavity models for arbitrarily shaped solutes which employ numerical solutions to the governing Poisson-Boltzmann equation [14–23]. These models and closely related model which employs an approximate solution to the Poisson-Bolzmann equation [24] have been applied to a variety of systems, establishing their efficacy in providing a useful description of electrostatic effects in solvation.



It is desirable to refine these models to obtain accurate results on the total solvation energy. There have been several problems in achieving this goal. Measurements of the transfer energies from solvents to vacuum [25–27] mix electrostatic contributions with contributions from other sources, notably from the solute-solvent nonpolar interactions and from the hydrophobic effect [28, 29]. In the current solvation models [24, 30, 31], these non-electrostatic contributions are assumed to be proportional to the solvent accessible surface area of the solute (SAS) [32, 33]. SAS is defined as the area swept out by the center of a solvent probe sphere rolled over the surface of the solute molecule [33]. Recently it was suggested, however, that the solute excluded volume can be a better fitting parameter than SAS [34, 35]. Parameterization of continuum solvent models is further exacerbated by significant sensitivity to the method used to represent the electronic state of the solute [18]. The implication of this sensitivity for molecular mechanics is that different force fields may require different parameterizations to obtain reliable accuracy. Recently an accurate continuum model was described in Ref. [31]. Both atomic charges and radii in this model were fitted to reproduce hydration free energies of the amino acid side chain and peptide backbone analogs, and related organic molecules. The dielectric constant of 2 was assigned to the solute interior to account for its polarizability. While this model provides accurate results for the listed solutes, it cannot be employed within molecular mechanics simulations since the current force fields are fitted to the vacuum dielectric constant of 1 and have their own atomic charges. A simple solvation model for use in molecular mechanical studies of peptides and proteins in solution was developed in Ref. [36] (see also [37]). This model treats all contributions to solvation free energy (including electrostatic

contributions) as proportional to SAS with different proportionality constants for different atom types obtained by fitting experimental data. Since the electrostatic energy generally is not proportional to SAS, transferability of such models should be tested for those compounds which were not used within the fitting procedure.

In this paper we report on the parameterization of a continuum solvation model for use in molecular mechanical studies based on the CFF91 force field [38]. The solvation model employs numerical solution to the Poisson-Bolzmann equation to describe electrostatic contributions to solvation while the non-electrostatic contributions are treated as proportional to SAS. Input to the parameterization process is the experimental solvation free energies for small organic molecules including amino acid analogs [25–27]. The results assign 20 independent variables. An agreement between the calculated and experimental free energies of solvation for 18 amino acid side-chain analogues is to within a mean unsigned error of 0.3 kcal/mol. To illustrate the application of the model to describe solvation effects on conformations, the Ramachandran map for the alanine dipeptide is computed and is found to be in good qualitative agreement with results from free energy simulations. Application of the model to 140 small molecule solutes shows a level of agreement with experiment which encourages its use in further molecular mechanical studies of solutes. Comparison to results reported from the solvation quantum-mechanical model AM1-SM2 [30] is provided.



## 2 METHODS

Electrostatic solvation free energies were computed using the continuum electrostatics program DelPhi [16, 17] developed at Columbia University and distributed by Biosym Technologies, Inc. This program solves the Poisson-Bolzmann equation using an optimal overrelaxation finite difference method. For the calculations reported in this paper, the salt concentration of the aqueous solvent was set to zero. Under this condition the Poisson-Bolzmann equation reduces to the well-studied Poisson equation from classical electrostatic theory [39]. Remaining input parameters to DelPhi consist of the solute-interior and solvent dielectric constants, solute atomic charges, and solute-atom and solvent-probe radii required to generate the molecular surface of the solute. The molecular surface used is that which is generated by the contact point between a spherical solvent probe and solute atoms as the probe rolls over the solute atoms [33]. This surface is used to separate the low-dielectric solute interior from the high-dielectric solvent.

Atomic partial charges were taken from the CFF91 force field [38]. The solvent dielectric constant was taken as 80. Consistency with evaluation of intra-molecular energies using the CFF91 force field demands the use of a solute internal dielectric constant of 1 (see Discussion). A solvent-probe radius of 1.4 A was used in generating SAS. In solving the Poisson equation for the solute-solvent system, the solute was first placed on a uniform $65 \times 65 \times 65$ cartesian grid. Atomic charges were then distributed onto grid points using an algorithm which preserves the dipole moment of the charge placement within its circumscribing grid box. It is worth noting that the dielectric constant in DelPhi is resolved to the midpoints of grid-lines connecting these grid-points, resulting in a finer description of the dielectric than of atomic charges. The closest distance from the molecular surface to the outer edge of the grid box was 5 A along each cartesian direction. While this procedure results

in variable grid resolution, the worst case for the solutes considered here corresponded to a grid spacing of 0.35 A.

The rigorous formulation of the electrostatic problem requires the Poisson-Boltzmann equation to satisfy the zero boundary condition at infinity [39]. A reliable approximation to the latter condition are the Coulomb boundary conditions [15] that are implemented in DelPhi. These boundary conditions are computed as the Coulomb potentials arising from solute charges in a uniform dielectric whose dielectric constant is that of solvent at outer edges of the grid box. The resulting potentials are then used as the Dirichlet boundary conditions for subsequent numerical calculations. The electrostatic solvation free energy was computed using the surface charge method [40, 41]. This method employs potentials on the cartesian grid to compute polarization charges residing on the boundary between the low-dielectric solute cavity and the high-dielectric solvent exterior. The electrostatic solvation free energy of the solute is given then as the Coulomb interaction energy between these charges and interior solute charges. An error is introduced into these calculations due to the representation of the solute surface on a cartesian grid. To reduce this error, the induced polarization charges are moved to ensure their correct placement [41]. This correction permits accurate results to be obtained using larger grid spacings. We estimate that errors in our calculated solvation free energies due to finite grid and due to the use of approximate boundary conditions are 5% or less. These estimates were obtained by using a larger number of the grid points (up to 128 × 128 × 128) and larger grid boxes.

The total solvation free energy $\Delta G$ is assumed to be the sum of an electrostatic contribution $\Delta Ge$ with nonpolar contributions $\Delta Gn$ from van der Waals interactions with solvent and from the hydrophobic effect [24, 29, 31]

$$\Delta G = \Delta Ge + \Delta Gn \tag{1}$$

$$\Delta Gn = \Delta G_{\text{hydrophobic}} + \Delta G_{\text{vdW}} \tag{2}$$

Indeed, solvation at constant pressure can be viewed in steps. The first step involves the introduction of a cavity into solvent. This step entails negligible work done against solvent pressure and a notable contribution from entropic reorganization of solvent. The latter effect captures the origins of the hydrophobic effect. In a second step the van der Waals interactions with solvent are turned on. These favorable interactions compensate the cost of cavity creation entailed in the first step. In a third step the electrostatic interactions with solvent are turned on. In our current model this last step can be addressed using the continuum electrostatic model described above.

A well-known property of linear alkanes is that their solvation free energies are roughly proportional to their SAS [28]. This observation suggests that the first two terms in this equation be combined into a single term $\Delta G_{\text{nonpolar}}$ described as proportional to SAS [24, 31]

$$\Delta G_n = \alpha 1 + \beta 1 \Sigma \tag{3}$$

where $\alpha 1$ and $\beta 1$ are parameters to be determined , $\Sigma$ is SAS. For the current work we adopt this form with SAS computed numerically using the Shrake-Rupley algorithm [32]. Note that SAS used to describe the nonpolar contributions to the solvation free energy is defined as the locus of the center of a solvent probe sphere rolled over the solute while the molecular surface used in the continuum electrostatic calculations is the locus of the contact point between the probe sphere and the solute

[33]. The same atomic radii are used to generate the molecular surface and SAS.

Recent discussions have focussed on the excluded volume effects in solvation phenomena [34, 35]. We have also explored the description of the nonpolar solvation energy as proportional to the excluded solute volume, V,

$$\Delta Gn = \alpha 2 + \beta 2V \tag{4}$$

Values of $V$ were computed using the method similar to that of [42].

In the complete model the total set of parameters to be determined consists of atomic radii, plus the appropriate linear proportionality constants $\alpha 1$ and $\beta 1$ or equivalently $\alpha 2$ and $\beta 2$. Given a training set of molecules whose experimental free energies of solvation are known, the problem of determing a set of parameters for this solvation model could be treated as a global optimization problem. However, the resulting values of these parameters can be rather sensitive to the types of molecules chosen for the training set. A case in point should illustrate the difficulties encountered. The Born expression for the electrostatic solvation free energy $\Delta G_{Born}$ in terms of the solvent dielectric constant $\epsilon$, the ion charge $q$, and the ion radius R [11]

$$\Delta G_{Born} = -q^2 / (2R)\ [1 - 1/\epsilon] \tag{5}$$

shows that the smaller the atomic radii are, the more negative (hydrophilic) is the solvation energy. The CFF91 atomic charges for a given radii set can yield "too negative" (in comparison with experimental data) energies for organic acids but "too positive" energies for ketones. Then the "optimal" radii are determined by the numbers of acids and ketones in the training set. Another concern is that atomic types used in the solvation model should be compatible with those employed in the CFF91 force field. Given the above considerations, the assignment of atomic radii was done by hand. Classes of molecules examined include linear and branched alkanes, cycloalkanes, alkenes, aromatic hydrocarbons, alcohols, acids, aldehydes, ketones, esters, amines, cyclic compounds containing H,C, and N, acetamide and other H,C,O, and N molecules, thiols and sulfides, bromine-, chlorine-, flourine-, and iodine-containing compounds, and phosphor-containing compounds. Due to their importance, amino acid side-chain analogs were included in the parameterization process explicitly. During the process, new atomic radii could be but were not necessarily introduced for each new atom type in the CFF91 force field. As new groups of molecules were considered, an inability to fit experimental data occassionally demanded that parameters already assigned be reexamined and new values applied over a broader range of molecules.

The parameter set resulting from this proceedure is not unique in the sense that it has not been developed as a global minimum-deviation fit to a selected set of compounds. The parameters were developed as a balance between the desire for a simple transferable set of parameters to describe the model, and coverage of as broad a range of molecular solutes as possible.

Calculations were performed on single conformations of the solutes which were obtained by energy minimization in vacuum using the CFF91 force field. When several minima existed for a given compound, the lowest energy conformation was selected in vacuum while the longest dipole moment conformation was selected in solutions. For investigations of the solvation free energy as a function of the dihedral angles of the alanine dipeptide, these angles were constrained to their desired values and the remainder of the molecule was permitted to relax using energy

minimization. All molecular mechanical manipulations were done using the programs INSIGHT and DISCOVER from Biosym Techologies, Inc..

## 3 RESULTS AND DISCUSSION

### 3.1 *Parametrization*

A logical starting point for our current parameterization are the linear alkanes whose electrostatic free energy contributes only a small fraction to the total free solvation energy (the CFF91 atomic charges dictate the electrostatic free energies of equal to 5–15% of the total solvation energies for the first ten alkanes). These compounds were used to assign values for the sp3 carbon and the hydrogen radii, and for fitting $\alpha 1$ and $\beta 1$ (and for $\alpha 2$ and $\beta 2$). Alkanes themselves provide a family of permitted values for the atomic radii which can effectively be compensated by a range of values for $\alpha 1$ and $\beta 1$ (or for $\alpha 2$ and $\beta 2$). However, the hydrogen radius of 1.5A and higher does not allow for fitting the solvation energies of polar molecules. The sp3 carbon and hydrogen radii were assigned 2A and 1A respectively similar to those chosen in the solvation model [31].

Then $\Delta G_{nonpolar}$ can be described using eq(3) with coefficients $\alpha 1 = 0.801 \pm 0.095$ kcal/mol and $\beta 1 = 0.0068 \pm 0.0003$ kcal/mol/A$^2$, or equivalently using eq(4) with coefficients $\alpha 2 = 1.16 \pm 0.08$ kcal/mol and $\beta 2 = 0.0040 \pm 0.0002$ kcal/mol/A$^3$. The difference between results using these two proportionalities does not exceed 0.2 kcal/mol for all 140 small organic molecules tested in this study. This equivalence between SAS and excluded volume proportionalities may reflect the fact that virtually all atoms in small molecules are exposed to water so that their excluded volumes and SAS are roughly proportional. SAS proportionality (3) is assumed futher. The final set of atomic radii for the current continuum solvation model is given in Table 1.

**Table 1** Atomic radii within the CFF91-based solvation model.

| *Atom* | *Potential type* | *Radius, A* |
|---|---|---|
| C | non-aromatic | 2.0 |
| | aromatic | 1.9 |
| H | | 1.0 |
| O | hydroxyl | 1.55 |
| | carbonyl | 1.3 |
| | ether or acetal | 1.0 |
| N | amine | 0.9 |
| | aromatic amine | 1.3 |
| | amide | 2.05 |
| | non-aromatic double bonded | 2.05 |
| | guandinium | 2.05 |
| | 3-membered ring | 1.4 |
| | 4-membered ring | 1.7 |
| | 5- or 6-membered ring | 1.6 |
| S | | 0.9 |
| Br | | 1.4 |
| Cl | | 1.8 |
| F | | 2.4 |
| I | | 0.6 |

### 3.2 *Solvation energies of small organic molecules*

The solvation free energies provided with the CFF91-based continuum model are given in Table 2 along with experimental values and with the quantum mechanical solvation model AM1-SM2 [30]. The accuracy for each group of compounds is discussed below.

HYDROCARBONS. The model provides very accurate results for linear and branched alkanes, alkenes, and simple aromatics. It does however, fail to forecast a decrease of solvation energies in a passage from the linear to the non-aromatic cyclic molecules with the same number of carbons.

H,C,O - COMPOUNDS. The model is accurate for alcohols, acids, ketones, esters, and non-aromatic aldehydes. The model provides an excess negative energy for aromatic aldehydes. On the other hand, energies of ethers, methoxy-compounds, and non-aromatic cyclo-compounds are "too hydrophobic": even the zero oxygen radius does not yield agreement with experiment when using the CFF91 charge set for these compounds.

H,C,N (,O) - COMPOUNDS. Besides the aromatics, the CFF91 charges are occasionally too small in magnitude to reproduce the hydrophilic nature of these molecules. However, the neutral analogues of amino acids are described quite accurately (see below and Table 3).

H,C,S - COMPOUNDS. The model describes accurately thiols and thiophenol but fails to describe molecules with C-S-C groups.

HALOGENS AND PHOSPHORUS - COMPOUNDS. Compounds considered are described well using the CFF91-based model.

In general the CFF91-based model achieves an accuracy comparable with that of the AM1-SM2 model [30]. Both models experience difficulties in describing some molecules (exclamation marks in Table 2 mark differences in the calculated and experimental data which exceed 1 kcal/mol). The CFF91 force field was originally developed with an emphasis on the amino acids, and the experimental solvation energies of their neutral analogues (excluding methionine; see Sec. 3.4) are reproduced quite well (the mean unsigned error is 0.3 kcal/mol; see Table 3). This allows for a description of the solvated peptides and proteins that is capable of capturing solvation effects on conformational preferences in solution. As an example, we apply the model to the solvated alanine dipeptide in the next section.

### 3.3 *Energetics of the solvated alanine dipeptide*

The alanine dipeptide (N-acetylalanyl-N-methylamide) has two major degrees of freedom, the Phi (C-N-C-C) and Psi (N-C-C-N) dihedral angles, which determine its conformation. Six basic conformations related to the conformational energy minima have been discussed in the literature [43-50]. These correspond to the following positions in (Phi,Psi) space: C7 axial (90, −90), C7 equatorial (−90, 90), C5 (−150, 150), alpha R (−70, −50), alpha L (70, 50), and PII (−80, 150). Experimental data confirm presence of C7 [51], alpha R, and PII [52] conformations in aqeous solvent.

The Ramachandran map for the alanine dipeptide in vacuum was computed using the CFF91 force field by fixing values for Phi and Psi on a 10 degree grid and

**Table 2** Calculated and experimental solvation free energies, kcal/mol.

| *Compound* | *Model* | | | | *Exp. [25-27]* |
|---|---|---|---|---|---|
| | *CFF91-based* | | *AM1-SM2 [30]* | | |
| | $\Delta Gn$ | $-\Delta Ge$ | $\Delta G$ | $\Delta G$ | $\Delta G$ |
| | | | HYDROCARBONS (35) | | |
| | | | Linear alkanes (10) | | |
| methane | 1.8 | 0.1 | 1.7 | - | 1.9 |
| ethane | 2.0 | 0.1 | 1.9 | 1.2 | 1.8 |
| propane | 2.2 | 0.2 | 2.0 | 1.4 | 2.0 |
| butane | 2.5 | 0.2 | 2.3 | 1.7 | 2.2 |
| pentane | 2.7 | 0.3 | 2.4 | - | 2.3 |
| hexane | 2.9 | 0.3 | 2.6 | 2.2 | 2.6 |
| heptane | 3.1 | 0.4 | 2.7 | 2.5 | 2.6 |
| octane | 3.3 | 0.4 | 2.9 | 2.7 | 2.9 |
| nonane | 3.5 | 0.5 | 3.0 | - | 3.0 |
| decane | 3.7 | 0.5 | 3.2 | - | 3.2 |
| Mean unsigned error | | | 0.06 | 0.4 | |
| | | | Chain alkanes (6) | | |
| 2-methylpropane | 2.4 | 0.2 | 2.2 | 1.6 | 2.3 |
| neopentane | 2.5 | 0.2 | 2.3 | 1.9 | 2.7 |
| 2-methylpentane | 2.8 | 0.3 | 2.5 | - | 2.5 |
| 2,4-dimethylpentane | 2.9 | 0.3 | 2.6 | 2.4 | 2.9 |
| 2,2,4-trimethylpentane | 3.1 | 0.3 | 2.7 | - | 2.9 |
| 2,2,5-trimethylhexane | 3.3 | 0.4 | 2.9 | - | 2.7 |
| Mean unsigned error | | | 0.2 | 0.7 | |
| | | | Cycloalkanes (5) | | |
| cyclopropane | 2.0 | 0.2 | 2.0 ! | 1.1 | 0.8 |
| cyclopentane | 2.4 | 0.3 | 2.1 | 1.6 | 1.2 |
| cyclohexane | 2.5 | 0.3 | 2.2 | 1.9 | 1.2 |
| methylcyclohexane | 2.7 | 0.3 | 2.4 | 2.1 | 1.7 |
| 1,2-dimethylcyclohexane | 2.8 | 0.3 | 2.5 | 2.7 | 2.9 |
| Mean unsigned error | | | 0.8 | 0.4 | |
| | | | Alkenes (6) | | |
| ethene | 2.0 | 1.0 | 1.0 | 0.8 | 1.3 |
| propene | 2.2 | 1.1 | 1.1 | 1.0 | 1.3 |
| 2-methylpropene | 2.4 | 0.8 | 1.6 | 1.1 | 1.2 |
| pentene | 2.6 | 1.2 | 1.4 | 1.4 | 1.3 |
| butadiene | 2.4 | 2.1 | 0.3 | 0.6 | 0.6 |
| cyclopentene | 2.4 | 1.1 | 1.3 | 1.0 | 0.6 |
| Mean unsigned error | | | 0.4 | 0.2 | |
| | | | Aromatic hydrocarbones (8) | | |
| benzene | 2.5 | 3.9 | −1.4 | −0.5 | −0.9 |
| toluene | 2.7 | 3.3 | −0.6 | −0.3 | −0.9 |
| o-xylene | 2.9 | 2.8 | 0.1 | −0.1 | −0.9 |
| p-xylene | 2.9 | 2.8 | 0.1 | 0.0 | −0.8 |
| naphtalene | 3.0 | 5.7 | −2.7 | −1.8 | −2.4 |
| anthracene | 3.4 | 7.4 | −4.0 | −2.5 | −4.2 |
| phenanthrene | 3.4 | 7.4 | −4.0 | - | −4.0 |
| ethylbenzene | 2.9 | 3.3 | −0.4 | - | -0.8 |
| Mean unsigned error | | | 0.5 | 0.8 | |
| Mean unsigned error over hydrocarbons | | | 0.3 | 0.5 | |

**Table 2** *Continued*

| *Compound* | *Model* | | | | *Exp. [25-27]* |
|---|---|---|---|---|---|
| | *CFF91-based* | | *AM1-SM2 [30]* | | |
| | *ΔGn* | *−ΔGe* | *ΔG* | *ΔG* | *ΔG* |
| | | | H,C,O - COMPOUNDS (36) | | |
| | | | Alcohols (7) | | |
| methanol | 1.9 | 7.3 | −5.4 | −5.8 | −5.1 |
| ethanol | 2.1 | 7.1 | −4.9 | −4.9 | −5.0 |
| 1-propanol | 2.3 | 7.1 | −4.8 | −4.6 | −4.8 |
| i-propanol | 2.3 | 6.9 | −4.6 | −4.1 | −4.8 |
| prop-2-en-1-ol | 2.3 | 7.9 | −5.6 | −4.9 | −5.0 |
| t-butanol | 2.5 | 6.6 | −4.1 | −3.3 ! | −4.5 |
| phenol | 2.6 | 10.1 | −7.5 | −5.8 | −6.6 |
| Mean unsigned error | | | 0.4 | 0.5 | |
| | | | Acids (3) | | |
| acetic acid | 2.1 | 9.5 | −7.4 | −7.7 | −6.7 |
| propinoic acid | 2.3 | 9.3 | −7.0 | −6.7 | −6.5 |
| butanoic acid | 2.5 | 9.3 | −6.8 | −6.3 | −6.4 |
| Mean unsigned error | | | 0.5 | 0.4 | |
| | | | Aldehydes (4) | | |
| acetaldehyde | 2.0 | 5.4 | −3.4 | −4.5 | −3.5 |
| propanal | 2.2 | 5.8 | −3.6 | −3.6 | −3.5 |
| butanal | 2.4 | 5.2 | −2.8 | −3.3 | −3.2 |
| pentanal | 2.7 | 6.0 | −3.3 | - | −3.0 |
| Mean unsigned error | | | 0.2 | 0.4 | |
| | | | Ketones (6) | | |
| propanone | 2.3 | 5.6 | −3.3 | −4.1 | −3.9 |
| butanone | 2.4 | 5.5 | −3.1 | −3.2 | −3.6 |
| 3-pentanone | 2.6 | 5.0 | −2.4 | −2.4 | −3.4 |
| 4-heptanone | 2.9 | 5.2 | −2.3 | −1.6 ! | −2.9 |
| 5-nonanone | 3.4 | 5.8 | −2.4 | −1.4 ! | −2.7 |
| acetophenone | 2.1 | 7.2 | −5.1 | −4.4 | −4.6 |
| Mean unsigned error | | | 0.6 | 0.7 | |
| | | | Esters (5) | | |
| methyl formate | 2.2 | 5.4 | −3.2 | −4.8 ! | −2.8 |
| methyl acetate | 2.4 | 5.1 | −2.7 | −4.0 | −3.3 |
| methyl propanoate | 2.7 | 5.5 | −2.8 | −3.0 | −2.9 |
| methyl butanoate | 3.0 | 5.1 | −2.1 | −2.7 | −2.8 |
| ethyl acetate | 2.6 | 5.1 | −2.5 | −3.3 | −3.1 |
| Mean unsigned error | | | 0.5 | 0.6 | |
| | | | Others (11) | | |
| dimethyl ether | 2.2 | 2.3 | −0.1 ! | −1.4 | −1.9 |
| diethyl ether | 2.5 | 2.5 | 0.0 ! | −0.2 ! | −1.6 |
| 1-methoxypropane | 2.2 | 1.9 | 0.3 ! | −0.5 ! | −1.7 |
| 2-methoxyethanol | 2.5 | 8.3 | −5.8 | −6.3 | −6.8 |
| 1,2-dimethoxyethane | 2.8 | 3.5 | −0.7 ! | −2.5 ! | −4.8 |
| tetrahydrofuran | 2.3 | 2.4 | −0.1 ! | −1.5 ! | −3.5 |
| 1,4-dioxane | 2.4 | 4.0 | −1.6 ! | −3.4 ! | −5.1 |
| 1,2-ethanediol | 2.2 | 13.1 | −10.9 ! | −9.6 ! | −7.7 |
| benzaldehyde | 2.7 | 8.3 | −5.6 ! | −4.7 | −4.0 |
| m-hydroxybenzaldehyde | 2.8 | 14.2 | −11.4 ! | −9.5 | −9.5 |
| p-hydroxybenzaldehyde | 2.8 | 14.4 | −11.6 ! | −9.8 | −10.5 |

**Table 2** *Continued*

| *Compound* | *CFF91-based* | | *Model* | | *Exp. [25-27]* |
|---|---|---|---|---|---|
| | *ΔGn* | *−ΔGe* | *ΔG* | *AM1-SM2 [30] ΔG* | *ΔG* |
| Mean unsigned error | | | 2.3 | 1.2 | |
| Mean unsigned error of H,C,O compounds | | | 1.0 | 0.8 (with Other) | |
| | | | 0.4 | 0.5 (without Others) | |
| | | | H,C,N - COMPOUNDS (20) | | |
| | | | Linear and chain molecules (7) | | |
| methanamine | 1.9 | 6.9 | −5.0 | - | −4.6 |
| ethylamine | 2.1 | 6.6 | −4.5 | −5.2 | −4.5 |
| 1-propylamine | 2.3 | 7.0 | −4.7 | −5.0 | −4.4 |
| 1-butanamine | 2.6 | 6.8 | −4.2 | −4.7 | −4.3 |
| dimethylamine | 2.2 | 2.9 | -0.7 ! | −4.3 | −4.3 |
| trimethylamine | 2.4 | 1.2 | 1.2 ! | −2.6 | −3.2 |
| propylguandine | 2.7 | 13.8 | −10.9 | - | −10.9 |
| Mean unsigned error | | | 1.3 | 0.5 | |
| | | | Cyclo-molecules (7) | | |
| aziridine | 2.1 | 7.6 | −5.5 | - | −5.4 |
| azetidine | 2.3 | 7.9 | −5.6 | −3.7 ! | −5.6 |
| pyrrolidine | 2.4 | 3.3 | -0.9 ! | −3.4 ! | −5.5 |
| piperazine | 2.4 | 6.1 | −3.7 ! | −7.8 | −7.4 |
| N-methylpiperazine | 2.7 | 3.8 | −1.1 ! | −6.2 ! | −7.8 |
| N,N'-dimethylpiperazine | 2.9 | 2.0 | 0.9 ! | −4.6 ! | −7,6 |
| methylimidazole | 2.5 | 12.6 | −10.1 | - | −10.3 |
| Mean unsigned error | | | 3.8 | 1.8 | |
| | | | Aromatic compounds (6) | | |
| aniline | 2.6 | 7.3 | −4.7 | −5.8 | −4.9 |
| pyridine | 2.4 | 7.2 | −4.8 | −4.4 | −4.7 |
| 4-methylpyridin | 2.6 | 6.6 | −4.0 | −4.2 | −4.9 |
| 2,6-dimethylpyridine | 2.9 | 5.4 | −2.5 ! | −2.3 ! | −4.6 |
| 2-methylpyrazine | 2.6 | 8.9 | −6.3 | −6.7 ! | −5.5 |
| methylindole | 3.0 | 9.9 | −6.9 | - | −5.9 |
| Mean unsigned error | | | 0.9 | 1.1 | |
| | | | H, C, O, N - COMPOUND S (5) | | |
| acetamide | 2.2 | 12.0 | −9.7 | - | −9.7 |
| propionamide | 2.4 | 11.3 | −8.9 | - | −9.4 |
| 2-methoxyethanamine | 2.5 | 8.7 | −6.2 | −7.1 | −6.6 |
| morpholine | 2.4 | 5.1 | −2.7 ! | −5.6 ! | −7.2 |
| 3-ethyl-2-methoxypyrazine | 3.1 | 8.8 | −5.7 ! | −5.8 ! | −4.4 |
| Mean unsigned error | | | 1.4 | 1.2 | |
| | | | H, C, S - COMPOUNDS (7) | | |
| methanethiol | 1.9 | 3.3 | −1.4 | −0.8 | −.1.2 |
| ethanethiol | 2.1 | 3.5 | −1.4 | −0.6 | −1.3 |
| thiophenol | 2.6 | 5.4 | −2.8 | −3.2 | −2.6 |
| dimethyl sulfide | 2.2 | 2.4 | −0.2 ! | −1.6 | −1.4 |
| methylethylsulfide | 2.5 | 2.4 | 0.1 ! | - | −1.5 |
| diethyl sulfide | 2.6 | 2.5 | 0.1 ! | −0.4 | −1.3 |
| thioanisole | 2.8 | 4.1 | −1.3 ! | −4.2 ! | −2.7 |
| Mean unsigned error | | | 0.8 | 0.7 | |
| | | | H, C, Br - COMPOUNDS (8) | | |
| bromomethane | 1.9 | 2.5 | −0.6 | −0.7 | −0.8 |
| bromoethane | 2.1 | 2.5 | −0.4 | −0.4 | −0.7 |

**Table 2** *Continued*

| *Compound* | *Model* | | | | *Exp. [25–27]* |
|---|---|---|---|---|---|
| | | *CFF91-based* | | *AM1-SM2 [30]* | |
| | *ΔGn* | *−ΔGe* | *ΔG* | *ΔG* | *ΔG* |
| 2-bromopropane | 2.3 | 2.3 | 0.0 | −0.2 | −0.5 |
| dibromomethane | 2.0 | 4.5 | −2.5 | −1.5 | −2.1 |
| 1,2-dibromoethane | 2.2 | 4.2 | −2.0 | −1.5 | −2.1 |
| bromobenzene | 2.6 | 3.2 | −0.6 | - | −1.5 |
| p-dibromobenzene | 2.7 | 2.7 | 0.0 ! | −3.2 | −2.3 |
| bromoform | 2.2 | 5.1 | −2.9 | −2.2 | −2.1 |
| Mean unsigned error | | | 0.7 | 0.4 | |
| | | | H, C, Cl - COMPOUNDS (8) | | |
| methyl chloride | 2.0 | 2.4 | −0.4 | −0.7 | -0.6 |
| dichloromethane | 2.2 | 4.2 | −2.0 | −1.2 | −1.4 |
| chloroform | 2.3 | 4.6 | −2.4 | −1.2 | −1.1 |
| ethyl chloride | 2.2 | 2.2 | 0.0 | −0.5 | −0.6 |
| 1,2-dichloroethane | 2.4 | 3.5 | −1.1 | −1.0 | −1.7 |
| 1,1,1-trichloroethane | 2.5 | 2.6 | −0.1 | −1.0 | −0.3 |
| 1,1,2-trichloroethane | 2.6 | 5.2 | −2.6 | −1.4 | −2.0 |
| chlorobenzene | 2.7 | 3.1 | −0.4 | −1.1 | −1.1 |
| Mean unsigned error | | | 0.6 | 0.3 | |
| | | | H, C, F - COMPOUNDS (4) | | |
| fluoromethane | 2.2 | 1.8 | 0.4 | 0.4 | -0.2 |
| trifluoromethane | 2.6 | 2.7 | −0.1 | 0.1 | 0.8 |
| tetrafluoromethane | 2.8 | 0.2 | 2.6 | 3.4 | 3.1 |
| 1,1-difluoroethane | 2.5 | 2.4 | 0.1 | 0.3 | -0.1 |
| Mean unsigned error | | | 0.6 | 0.5 | |
| | | | H, C, I - COMPOUNDS (4) | | |
| iodomethane | 1.8 | 2.8 | −1.0 | −1.1 | −0.9 |
| iodoethane | 2.0 | 2.8 | −0.8 | −0.7 | −0.7 |
| 1-iodopropane | 2.3 | 2.7 | −0.4 | −0.5 | −0.6 |
| 1-iodobutane | 2.5 | 2.8 | −0.3 | −0.2 | −0.3 |
| Mean unsigned error | | | 0.1 | 0.1 | |
| | | | COMPOUNDS WITH 4 KINDS OF ATOMS INCLUDING HALOGENES (9) | | |
| bis-2-chloroethylsulfide | 3.0 | 5.8 | −2.8 ! | −3.2 | −3.9 |
| 2-bromo-1-chloroethane | 2.3 | 3.9 | −1.6 | −1.3 | −2.0 |
| chlorofluoromethane | 2.3 | 3.7 | −1.4 | −0.7 | −0.8 |
| chlorodifluoromethane | 2.5 | 3.4 | −0.9 | −0.4 | −0.5 |
| 2-chloro-1,1,1-trifluoro ethane | 2.8 | 3.3 | −0.5 | 0.5 | 0.1 |
| bromotrifluoromethane | 2.7 | 0.7 | 2.0 | 0.5 ! | 1.8 |
| 2,2,2-trifluoroethanol | 2.8 | 8.0 | −5.2 | −4.5 | −4.3 |
| 1,1,1-trifluoropropan-2-ol | 2.9 | 7.4 | −4.5 | −3.2 | −4.2 |
| p-bromophenol | 2.7 | 9.4 | −6.7 | −7.0 | −7.1 |
| Mean unsigned error | | | 0.5 | 0.5 | |
| | | | COMPOUNDS CONTAINING P (4) | | |
| phosphine | 1.7 | 0.2 | 1.5 | 0.6 | 0.6 |
| P (OCH3) 3 | 2.8 | 11.2 | −8.4 | −6.7 ! | −8.7 |
| P (OC2H5) 3 | 3.4 | 11.5 | −8.1 | −4.4 ! | −7.8 |
| P (OC3H7) 3 | 4.1 | 11.2 | −7.1 | −3.2 ! | −6.1 |
| Mean unsigned error | | | 0.6 | 2.1 | |

Total number of molecules: 140

**Table 3** Solvation energies of the neutral amino acid analogues, kcal/mol.

| *Amino acid* | *Molecule* | *Theory* *CFF91* | *AM1-SM2 [30]* | *Exp. [27]* |
|---|---|---|---|---|
| Ala | methane | 1.7 | - | 1.9 |
| Arg | propylguandine | −10.9 | - | −10.9 |
| Asn | acetamide | −9.7 | - | −9.7 |
| Asp | acetic acid | −7.4 | −7.7 | −7.6 |
| Cys | methanethiol | −1.4 | −0.8 | −1.2 |
| Gln | propionamide | −8.9 | - | −9.4 |
| Glu | propionic acid | −7.0 | −6.7 | −6.5 |
| His | methylimidazole | −10.1 | | −10.2 |
| Ile | butane | 2.3 | 1.7 | 2.2 |
| Leu | 2-methylpropane | 2.2 | 1.6 | 2.3 |
| Lys | butanamine | −4.2 | −4.7 | −4.3 |
| Met | methylethylsulfide | 0.1 | - | −1.5 |
| Phe | toluene | −0.6 | −0.3 | −0.9 |
| Ser | methanol | −5.4 | −5.8 | −5.1 |
| Thr | ethanol | −4.9 | −4.9 | −4.9 |
| Trp | methylindole | −6.9 | - | −5.9 |
| Tyr | p-cresole | −6.8 | - | −6.1 |
| Val | propane | 2.0 | 1.4 | 2.0 |
| Mean unsigned error | | 0.3 | 0.4 | |



subjecting the all other degrees of freedom to the energy minimization in the program DISCOVER [38]. The results shown graphically in Figure 1 and numerically in Table 4 evidence three conformational minima, C5, C7 axial and C7 equatorial, whose locations are consistent with those found in other molecular mechanical [43–48] and quantum mechanical studies [49, 50].

Our solvation model was applied to each minimized conformation on the grid. The nonpolar component of the solvation energy which proves to be insensitive to the dipeptide conformational changes ($\Delta Gn = 3.0 \pm 0.2$ kcal/mol). The results on the electostatic component of the solvation energy varied over a considerably larger range (Figure 2), exploring the expanded interval $\Delta Ge = -(12\text{–}24)$ kcal/mol. The Ramachandran map for the total solvated conformational energy of the solvated alanine dipeptide, computed as the sum of the CFF91 force field and the solvation energy produced by our current model, is given in Figure 3 (See also Table 4). This map shows the global minimum shifted away from the C7 equatorial conformation in vacuum to a broad basin in the beta sheet region of the map which spans nearly 45 degrees in Phi and Psi and which subsumes the PII and C5 conformations. Two additional minimum energy conformations, corresponding to alpha L and alpha R conformations, appear due to solvation. The relative stabilization of alpha L, alpha R, and PII due to solvation effects has been reported earlier by Pettit and Karplus in an integral equation studies [45]. Their results retained however the global minimum at the C7 equatorial conformation, that may reflect the differences in the CFF91 force field and that of used in Ref. [45]. The shift due to solvation effects of the global minimum away from the C7 equatorial conformation is also in qualitative agreement with results from Monte Carlo free energy simulations reported by Mezei *et al.* [44], which demonstrated preferential stabilization of the PII and alpha R conformations. Our results are in remarkably good agreement with those obtained using molecular dynamics free-energy simulations by Anderson and

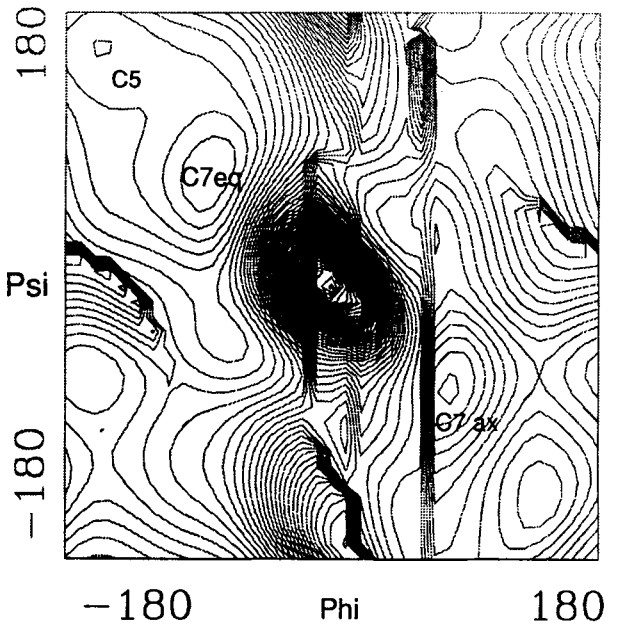


**Figure 1** The CFF91 force field energy surface of the alanine dipeptide in vacuum (energy is contoured at 1 kcal/mol intervals).

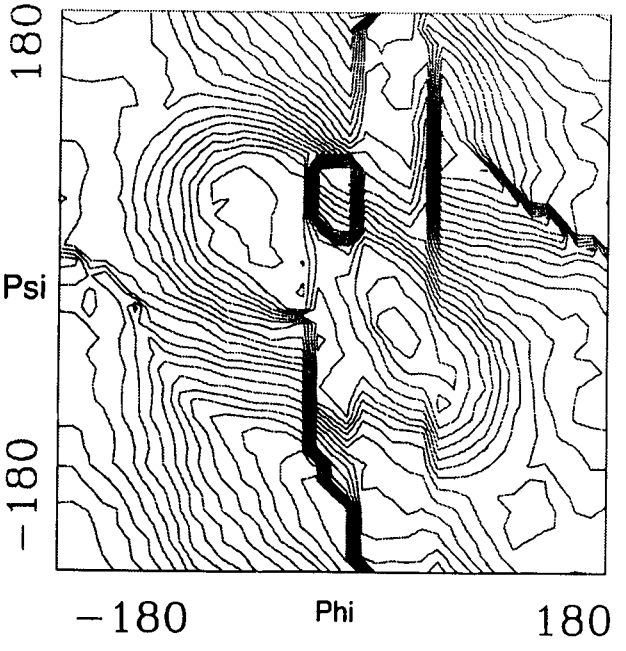


**Figure 2** The CFF91-based solvation electrostatic energy surface of the alanine dipeptide (energy is contoured at 1 kcal/mol intervals).

**Table 4** CFF91-based minima of the alanine dipeptide energy.

| *Conformation* | *Vacuum* | | | *Solvated* | | | *Solvation* |
|---|---|---|---|---|---|---|---|
| | *Phi* | *Psi* | *Energy* | *Phi* | *Psi* | *Energy* | *Energy* |
| C5 | −150 | 160 | 0 | −150 | 150 | 0 | 0 |
| C7 axial | 80 | −70 | −0.3 | 80 | −70 | 3.2 | 2.9 |
| C7 equatorial | −80 | 80 | −2.0 | - | - | - | - |
| alpha L | - | - | - | 70 | 70 | 1.6 | −6.3 |
| alpha R | - | - | - | −90 | −60 | 0.9 | −3.3 |
| PII | - | - | - | −70 | 150 | −0.4 | −1.3 |

Relative to *C5* conformation in kcal/mol.

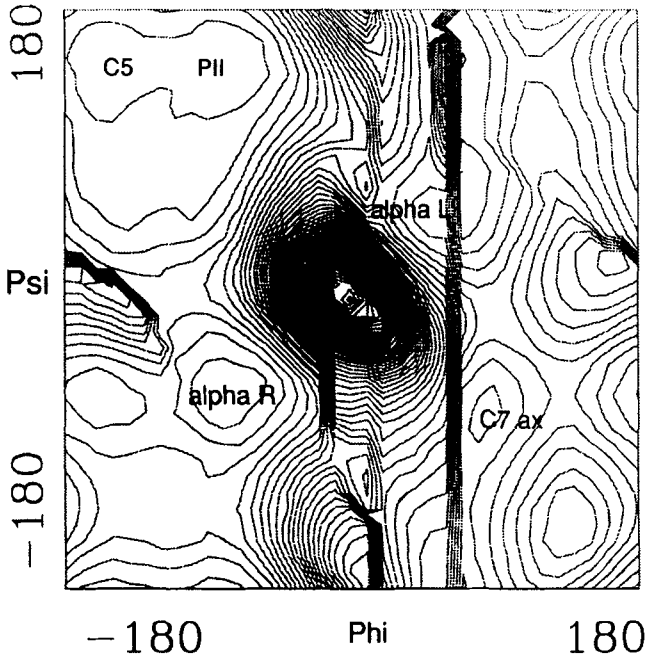


**Figure 3** The CFF91-based conformational energy surface of the solvated alanine dipeptide (energy is contoured at 1 kcal/mol intervals).

Hermans [46]. The distinctive feature of the map obtained by authors is a broad global free-energy minimum centered on the beta confomation (Phi = −110, Psi = 120), which corresponds well to the large low-energy basin evidenced in Figure 3.

We have also computed the Ramachandran map for the solvation free energy of the alanine dipeptide using the Wesson-Eisenberg model [36] (see Figure 4). While results are qualitatively similar to those shown in Figure 2, they evidence a dramatically narrower range of variation in solvation energy $\Delta Ge = -(7\text{-}11)$ kcal/mol.

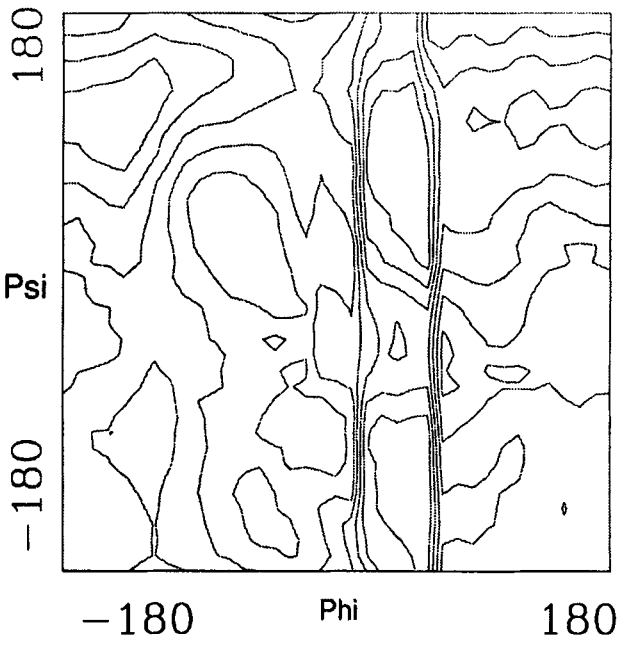


**Figure 4** The solvation energy surface of the alanine dipeptide from the Wesson-Eisenberg model [36] (energy is contoured at 0.5 kcal/mol intervals).



### 3.4. *Directions for improvement to the current model*

Our current model does not account for solute polarizability. This physical property appears in a natural way when the quantum mechanical description of a solute is employed. The current force fields include the solute polarizability only implicitly since the atomic partial charges are derived to yield an appropriate description of inter-molecular forces in condensed phase. Hence, any inclusion of solute polarization effects requires a concomitant change in the vacuum charges assigned to solutes. While solute polarization can be incorporated into our continuum model through the introduction of an internal dielectric constant [18], we have chosen to sacrifice this effect to achieve internal consistency with the use of the CFF91 force field charges to describe intra- and inter-molecular solute interactions. It should be noted, however, that the CFF91 atomic charges may be themselves a source of errors. In particular, we have mentioned that the current model overestimates the solvation free energies of cycloalkanes. The requirement of transferability of the force field parameters would strongly bias its development towards adopting the same atomic partial charges for carbons in linear and cyclic alkanes. Recent quantum mechanical studies show that this transferability might be inadequate especially for short cycloalkanes though results are strongly dependent on the method used to extract atomic partial charges from the electron density [53, 54].

An exception to the achieved accuracy of the model for amino acid analogues (see Table 3) is methylethylsulfide (analogue of the methionine side chain). Given the CFF91 partial charge assigned to sulfur, even the zero sulfur radius fails to

replicate the hydrophilic nature of this compound. A possible cause of this discrepancy may be related to the high polarizability of sulfur. In the quantum mechanical study [55], it was shown that the fitting the electrostatic potential of the sulfur-containing amino acids requires to introduce the dummy atomic charges. These comments may also apply to the relatively poorer agreement with experiment for C-S-C compounds in Table 2.

Another place for improvement in our solvation model is its simple treatment of nonpolar contributions to solvation free energy. By assuming that all SAS is equal, an implicit assumption is made that the van der Waals interactions between the solute and solvent atoms average to a uniform representation over SAS. It has been suggested that the hydrophobic effect should depend on the shape of the solute surface [56–58]. Our current model would anticipate an increase in the hydrophobic cost of solvation on the addition of a methyl group to any molecule due to the increase of SAS. The experimental data do not show this systematic increase. A curvature-sensitive representation of the hydrophobic effect similar to that suggested in Ref. [58] would reduce the hydrophobic cost for the introduction of a methyl group relative to larger-radius heavy atoms due to a decrease in the local radius of curvature around the hydrogen atoms. Curvature effects may similarly play a role in providing a distinction between the linear and cyclic alkane solvation.

The methods described in this paper are appropriate for computing the conformationally-dependent solvation free energy of a molecular solute in aqeous solution. In its presented form it can be used to compare quite different conformations and is immediatelly applicable to the Monte Carlo conformational search methods to locate low-energy conformers. The incorporation of the model into minimization and molecular dynamics algorithms demands the evaluation of derivatives of the solvation free energy with respect to changes in atom coordinates. While we have not described evaluation methods here, adequate computational machinery has been developed and is described elsewhere [59]. A word of caution on the applicability of continuum models in molecular mechanics may be in order. Continuum methods describe ensemble-averaged behaviors of solvent. As such, they are well-matched to the relative ranking of conformations in solution as is done here, and to Monte Carlo methods whose goal is the generation of ensembles of solvated conformations. However, time steps in molecular dynamics are typically on the order of 1.E-15 seconds. Dielectric relaxation times associated with the response of water to changes in the location of solute charges and dielectric boundaries are orders of magnitude larger [40]. When continuum methods are used in molecular dynamics simulations, these long time-scale responses are instantaneously updated as short time-scale motions are explored. Thus the effects of solvent on short time-scale dynamical properties of the system will not be well-described. The use of molecular dynamics to generate an ensemble of conformations similar to that which would be generated using Monte Carlo may however inherit an appropriately ensemble-averaged solvent representation using these models.



## 4 CONCLUSIONS

We have presented the solvation model for use in molecular mechanical computations employing a particular choice of force field, CFF91, which should enable simulations of molecules in solution. Results obtained for 140 solute molecules

replicate the solvation free energies of these molecules to within approximately the level of accuracy as that achieved using the quantum mechanical model AM1-SM2. In particular accurate results are obtained for amino acid neutral analogues (mean unsigned error of 0.3 kcal/mol). The results of application of the model to the alanine dipeptide in solution shows that characteristic features obtained using more computationally intensive methods to examine the effects of solvent on conformational preferences can be replicated using the current model.

*Acknowledgements*

The authors would like to gratefully acknowledge fruitful discussions with Barry Honig and Doree Sitkoff of Columbia University, with Arnie Hagler and Judy Hemple of Biosym, with Frank Brown of Glaxo Inc, and with Alex Rashin.

*References*

[1] A. Warshel and S. Russel, "Calculation of electrostatic interactions in biological systems and in solution", *Quart. Rev. Biophys.* **17**, 283 (1984).
[2] D.L. Beveridge and F.M. DiCapua, "Free energy via molecular simulation: applications to chemical and biomolecular systems", *Annu. Rev. Biophys. Chem.* **18**, 431 (1989).
[3] T.P. Straatsma and J.A. McCammon, "Computational Alchemy", *Annu. Rev. Phys. Chem.* **43**, 407 (1992).
[4] A.A. Rashin, "Aspects of protein energetics and dynamics", *Progr. Biophys. Mol. Biol.* **60**, 73 (1993).
[5] S. Miertus, E. Scrocco, and J. Tomasi, "Electrostatic interaction of a solute with continuum. A direct utilitization of ab initio molecular potentials for the prevision of solvent effects", *Chem. Phys.* **55**, 117, (1981).
[6] M.J. Field, P.A. Bash, and M. Karplus, "A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations", *J. Comput. Chem.* **11**, 700 (1990).
[7] J. Gao, F.J. Luque, M. Orozco, "Induced dipole moment and atomic charges based on average electrostatic potentials in aqueous solutions", *J. Chem. Phys.* **98**, 2975 (1993).
[8] G.P. Ford and B. Wang, "Incorporation of hydration effects within the semiempirical molecular orbital framework. AM1 and MNDO results for neutral molecules, cations, anions, and reacting systems", *J. Amer. Chem. Soc.* **114**, 10563 (1992).
[9] K. Baldridge, R. Fine, and A. Hagler, *J. Comput. Chem.* (in press).
[10] A. Rashin, M. Bukatin, J. Andzelm, and A. Hagler, *Biophysical Chemistry* (in press).
[11] M. Bom, "Volumes and heats of hydration of ions", *Z. Phys.* **1**, 45 (1920).
[12] C. Tanford and J.G. Kirkwood, "Theory of protein titration curves. I. General equations for impenetrable spheres", *J. Amer. Chem. Soc.* **79**, 5333 (1957).
[13] M.K. Gilson, A. Rashin, R. Fine and B. Honig, "On the calculation of electrostatic interactions in proteins", *J. Mot. Biol.* **183**, 503 (1985).
[14] J. Warwicker and H.C. Watson, "Calculation of the electric potential in the active cleft due to α-Helix dipoles", *J. Mol. Biol.* **157**, 671 (1982).
[15] I. Klapper, R. Hangstrom, R. Fine, K. Sharp, and B. Honig, "Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino acid modification", *Proteins* **1**, 47 (1986).
[16] M.K. Gilson, K.A. Sharp, and B. Honig, "Calculating the electrostatic potential of molecules in solution: Method and errors assessment" *J. Comp. Chem.* **9**, 327 (1987).
[17] A. Nicholls and B. Honig, "A Rapid Finite Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson-Boltzmann Equation", *J. Comp. Chem.* **12**, 435 (1991).
[18] K. Sharp, A. Jean-Charles and B. Honig, "A local dielectric constant model for solvation free energies which accounts for solute polarizability", *J. Phys. Chem.* **96**, 3822 (1992).
[19] M.E. Davis and J.A. McCammon, "Electrostatics in Biomolecular Structure and Dynamics", *Chem. Rev.* **10**, 386 (1990).
[20] A.A. Rashin, "Hydration phenomena, classical electrostatics, and the boundary element method", *J. Phys. Chem.* **93**, 4664 (1989).
[21] A.H. Juffer, E.F.F. Botta, B.A.M. van Keulen, A. van der Ploeg, and H.J.C. Berendsen, "The

electric potential of a macromolecule in a solvent: a fundamental approach", *J. Comp. Phys.* **97**, 144 (1991).
[22] R.J. Zauhar and R.S. Morgan, "The rigorous computation of the molecular electric potential", *J. Comp. Chem.* **9**, 171 (1988).
[23] T.J. You and S.C. Harvey, "Finite element approach to the electrostatics of the macromolecules with arbitrary geometries", *J. Comp. Chem.* **14**, 484 (1993).
[24] W.C. Still, A. Tempczyk, R.C. Hawley and T. Hendrickson, "Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Amer. Chem. Soc.* **112**, 6129 (1990).
[25] S. Cabani, P. Gianni, V. Mollica and L. Lepori, "Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution", *J. Solution Chem.* **10**, 563 (1981).
[26] A. Ben-Naim and Y. Marcus, "Solvation thermodynamics of nonionic solutions", *J. Chem. Phys.* **81**, 2016 (1984).
[27] A. Radzicka and R. Wolfenden, "Comparing the polarities of the amino acids: side chain distribution coefficients betweenthe vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution', *Biochemistry* **27**, 1664 (1988).
[28] C.H. Tanford, The Hydrophobic Effect, Wiley, NY, 1980.
[29] A. Ben-Naim, Solvation Thermodynamics, Plenum Press, NY, 1987.
[30] C.J. Cramer and D.G. Truhlar, "AM1-SM2 and PM3-SM3 parametrized SCF solvation models for free energies in aqueous solution", *J. Comput. -Aided Mol. Des.* **6**, 629 (1992).
[31] D. Sitkoff, K.A. Sharp, and B. Honig, "Accurate calculation of hydration free energies using macroscopic solvent models", *J. Phys. Chem.* **98**, 1978 (1994).
[32] A. Shrake and J.A. Rupley, "Environment and exposure to solvent of protein atoms. Lisozyme and insuline", *J. Molec. Biol.* **79**, 351 (1973).
[33] F.M. Richards, "Areas, volumes, packing, and protein structure", *Ann. Rev. Biophys. Bioeng.* **6**, 151 (1977).
[34] A. Ben-Naim and R.M. Mazo, "Size dependence of the solvation energies of large solutes", *J. Phys. Chem.* **97**, 10829 (1993).
[35] A.B. Schmidt, "Excluded volume effect in solvation thermodynamics: simple Lennard-Jones model", *Biophys Chem* (in press).
[36] L. Wesson and D. Eisenberg, "Atomic solvation parameters applied to molecular dynamics of proteins in solution", *Protein Science* **1**, 227 (1992).
[37] C.A. Schiffer, J.W. Caldwell, P.A. Collman, and R.M. Stroud, "Protein structure prediction with a combined solvation free energy-molecular mechanics force field", *Molecular Simulaton* **10**, 121 (1993).
[38] DISCOVER User Guide, Biosym Technologies, 1993, San Diego, CA 92121.
[39] J.D. Jackson, Classical Electrodynamics, Wiley, NY, 1962.
[40] C.J.F. Bottcher, Theory of Electric Polarization, 2nd ed., Elsevier, NY, 1973.
[41] A. Nicholls and B. Honig (unpublished results)
[42] A.Y. Meyer, "Molecular mechanics and molecular shape. V. On the computation of the bare surface area of molecules", *J. Comp. Chem.* **9**, 18 (1988).
[43] A.T. Hagler, D.J. Osguthorpe, and B. Robson, "Monte Carlo simulation of water behavior around the dipeptide N-acetylalanyl-N-nethylamide", *Science* **208**, 599 (1980).
[44] M. Mezei, P.K. Mehotra, and D.L. Beveridge, "Monte Carlo determination of the free energy and internal energy of hydration for Ala dipeptide at 25 C", *J. Amer. Chem. Soc.* **107**, 2239 (1985).
[45] B.M. Pettitt and M. Karplus, "The potential of mean force surface for the alanine dipeptide in aqueous solution: a theoretical approach", *Chem. Phys. Lett.* **121**, 94 (1985).
[46] A.G. Anderson and J. Hermans, "Microfolding: conformational probability map for the alanine dipeptide in water from molecular dynamics simulations", *Proteins* **3**, 262 (1988).
[47] T. Lazaridis, D.J. Tobias, C.L. Brooks III, and M.E. Paulaitis, "Reaction paths and free energy profiles for conformational transitions", *J. Chem. Phys.* **95**, 7612 (1991).
[48] P.E. Smith, B.M. Pettitt, and M. Karplus, "Stochastic dynamics simulations of the alanine dipeptide using a solvent-modified potentialenergy surface", *J. Phys. Chem.* **97**, 6907 (1993).
[49] T. Head-Gordon, M. Head-Gordon, M.J. Frisch, C.L. Brooks III, and J.A. Pople, "Theoretical study of blocked glycine and alanine peptide analogues", *J. Amer. Chem. Soc.* **113**, 5989 (1991).
[50] H.-J. Bohm and S. Brode, "Ab initio SCF calculations on low-energy conformers of N-acetyl-N-methylalaninamide and N-acetyl-N-methylglycinamide", *J. Amer. Chem. Soc.* **113**, 7129 (1991).
[51] M. Avignon, C. Garrigou-Lagrange, and P. Bothorel, "Conformational analysis of dipeptides in aqueous solution. II. Molecular structure of lycine and alanine dipeptides by depolarized Rayleigh scattering and laser Raman spectroscopy", *Biopolymers* **12**, 1651 (1973).

[52] V. Madison and K.D. Kopple, "Solvent-dependent conformational distributions of some peptides" *J. Amer. Chem. Soc.* **102**, 4855 (1980).
[53] K.B. Wiberg and P.R. Rablen, "Comparison of atomic charges derived via different procedures", *J. Comp. Chem.* **14**, 1504 (1993).
[54] C. Chipot, J.G. Angyan, G.G. Ferenczy, and H.A. Scheraga, "Transferable net atomic charges from a distributed multipole analysis for the description of electrostatic properties. A case study of saturated hydrocarbons", *J. Phys. Chem.* **97**, 6628 (1993).
[55] C. Chipot, B. Maigret, J.-L. Rivail, and H.A. Scheraga, "Modelling amino acid side chains. 1. Determination of net atomic charges from ab initio self-consistent-field molecular electrostatic properties", *J. Phys. Chem.*, 1992, 96, 10276. *J. Phys. Chem.* **96**, 10276 (1992).
[56] R.B. Hermann, "Use of solvent cavity area and number of packed solvent molecules around a solute in regard to hydrocarbon solubility and hydrophobic interactions", *Proc. Natl. Acad. Sci. USA* **74**, 4144 (1977).
[57] K.A. Sharp, A. Nicholls, R.M. Fine, and B. Honig, "Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects", *Science* **252**, 106 (1992).
[58] A. Nicholls, K.A. Sharp, and B. Honig, "Protein folding and association: insights from interfacial and thermodynamic properties of hydrocarbones", *Proteins* **11**, 281 (1991).
[59] M. Gilson, M. Davis, B. Luty, and J.A. McCammon, "Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation", *J. Phys. Chem.* **97**, 3591 (1993).