

# Quantum mechanical geometry optimization in solution using a finite element continuum electrostatics method

Christian M. Cortis<sup>a)</sup>

Department of Applied Physics and Center for Biomolecular Simulation, Columbia University,  
New York, New York 10027

Jean-Marc Langlois

Schrödinger, Inc., 121 SW Morrison, Suite 1212, Portland, Oregon 97204

Michael D. Beachy and Richard A. Friesner<sup>b)</sup>

Department of Chemistry, and Center for Biomolecular Simulation, Columbia University,  
New York, New York 10027

(Received 19 February 1996; accepted 22 May 1996)

We present a new algorithm for performing *ab initio* solution phase geometry optimizations. The procedure is based on the self consistent-reaction-field method developed in our laboratory which combines electronic structure calculations with a finite element formulation of the continuum electrostatics problem. A gradient for the total solution phase free energy is obtained by combining different contributions from the gradient of the classical polarization free energy and the derivatives of the quantum mechanical energy. The method used in obtaining the classical gradient is based on exact linear algebra relations and a Green function formalism due to Handy and Schaefer. Both the classical and quantum mechanical gradients are validated by comparison with energy finite differences. The result of applications to a number of small organic compounds are discussed. Comparisons between the predicted location and depth of the various solution phase minima of the Ramachandran map for the alanine dipeptide and those reported by Gould *et al.* are also presented.

© 1996 American Institute of Physics. [S0021-9606(96)01533-4]

## I. INTRODUCTION

In recent years, there has been a great deal of effort devoted to the development of self-consistent-reaction-field (SCRf) methods for the calculation of molecular solvation free energies. We and others,<sup>2-4</sup> have employed *ab initio* correlated wave functions, in conjunction with solvent dielectric continuum models, to carry out these calculations. With appropriate optimization of the atomic radii in the continuum model, reasonable agreement with experiment has been achieved.<sup>5</sup> In our most recent calculations, using a test suite of 120 small molecules, an average difference of 0.75 kcal/mol with experimental results is observed. Also, we have investigated systematic errors in dielectric continuum solvation models and developed a correction scheme based upon first shell hydrogen bonding corrections,<sup>6</sup> reducing the average error by approximately a factor of 2.

In typical SCRf calculations reported to date, the molecular geometry has either been obtained from molecular mechanics force fields or gas phase energy minimization. For small, relatively rigid molecules, this is a good approximation to the solution phase geometry. However, as one studies larger and more flexible molecular structures, significant changes in the geometry due to solvation effects can be expected. Consequently, it is essential for such applications to

develop a gradient methodology for solution phase geometry optimization of SCRf calculations.

A number of methods have been presented in the literature for carrying out quantum chemical solution phase geometry optimization. Some of these are based upon simplified models for the dielectric cavity of the solute (spherical, ellipsoidal) sometimes combined with drastic approximations to the solute charge density, such as a multipole series truncated at low order.<sup>7-9</sup> Our investigations indicate that these sorts of approximations are inadequate to yield an accuracy of 1 kcal/mol or better, which our current parameterization appears to be capable of achieving. Consequently, we focus on approaches which treat the full solute charge distribution in a realistic molecular cavity.

Work on a model of this type has been carried out by Tomasi and co-workers,<sup>10,11</sup> and has been implemented in the GAUSSIAN 94 suite of *ab initio* electronic structure programs. However, at present there are no results of this methodology available for large molecules, where solution phase geometry optimization is critical to obtaining accurate results. The presentation of a novel algorithmic formalism for solution phase SCRf geometry optimization in the context of the PS-GVB suite of electronic structure programs, followed by a significant number of nontrivial molecular applications, is the objective of the present paper.

The paper is organized as follows. In Sec. II, we give a brief description of the SCRf methodology implemented in PS-GVB, including the hydrogen bonding corrections presented by Marten *et al.*<sup>6</sup> We then discuss in Sec. III, our novel finite element numerical methodology for solving the

<sup>a)</sup>Present address: Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032; tel: 212-305-6884; electronic address: chris@boreale.bioc.columbia.edu

<sup>b)</sup>Corresponding author. tel: 212-854-7006, electronic address: rich@ccubs.chem.columbia.edu

Poisson–Boltzmann (PB) equation: this will be described in detail in another publication,<sup>12</sup> so what is presented here is a relatively brief outline. In Sec. IV, a method for calculating the analytical gradient of the PB energy is discussed and compared with finite difference calculations to establish its accuracy. Section V then describes the full SCRf gradient methodology and a variety of test cases, including a merocyanine dye molecule. Results for the alanine dipeptide are presented in Sec. VI. (Finally, in the conclusion, we discuss future directions of this research.)

## II. OVERVIEW OF SCRf METHODOLOGY

As a mean-field theory of solvent response to the solute electrostatic field, the dielectric continuum model eliminates the need for averaging over solvent configurations in solvation calculations. Because of this powerful feature, the model has proven to be especially useful in biophysical applications and molecular modeling, even when applied to parametrized representations of the molecular charge distributions. (The reader is referred to a review by Sharp and Honig for a detailed discussion of the features of the model and its applications.<sup>13</sup> Another review by Tomasi and Persico<sup>14</sup> provides additional details regarding applications to *ab initio* calculations.) To assess the model's limitations however, it has been necessary to couple it to the most accurate representation available of solute charge distributions. This has been the central principle behind the development of all SCRf methods currently in use. These methods in general allow a quantum mechanical representation of a solute to be coupled to a dielectric continuum model of the solvent, thus enabling *ab initio* solution phase calculations to be carried out. The effect of the solvent continuum is represented by an induced surface charge distribution located at the solute–solvent dielectric boundary. The total quantum mechanical energy of the system can be written as

$$E_s^{QM} = \langle \psi^s | H^0 | \psi^s \rangle + \frac{1}{2} [\langle \psi^s | H' | \psi^s \rangle + H''], \quad (2.1)$$

where  $H^0$  is the gas phase Hamiltonian,  $H'$  is the one-electron interaction with the reaction field,  $H''$  is the nuclear interaction with the reaction field, and  $\psi^s$  is the solution phase wave function.

In the specific implementation discussed by Tannor *et al.*<sup>5</sup> and Marten *et al.*<sup>6</sup> the quantum mechanical charge density obtained from a solution to the gas phase or “free molecule” problem ( $H' = H'' = 0$ ) is represented by a set of point charges centered on the atoms. The magnitude of the charges is calculated by fitting the Coulomb potential they produce on a grid some finite distance from the molecule, to the full molecular electrostatic potential. This process is usually referred to as electrostatic potential (ESP) fitting.

The point charges are then used to solve the dielectric continuum problem, which will be discussed below. The source of the reaction field is represented as an induced surface polarization charge and is used to generate a first iteration of the  $H'$ ,  $H''$  terms in the Hamiltonian. The modified quantum mechanical problem is solved to produce a new molecular charge density and the process is repeated until

convergence. In this framework, the solvation free energy is then given by the sum of electrostatic, nonpolar, and hydrogen bonding terms:

$$\Delta G^f = \Delta G^{es} + \Delta G^{np} + \Delta G^{sa}. \quad (2.2)$$

In our calculations, the model used to evaluate the nonpolar contribution  $\Delta G^{np}$  is based on experimentally determined vacuum to water solvation energies of linear and branched alkanes<sup>15</sup> and is linear in the solvent accessible surface area  $A$ .

The need for a correction term to account for hydrogen bonding effects has recently been established in our laboratory.<sup>6</sup> The evidence is based on the observation that the discrepancy between experimental and predicted solvation free energies using  $\Delta G^{es} + \Delta G^{np}$  instead of Eq. (2.2), for certain compounds is correlated with the strength of hydrogen bonds between some functional groups and the solvent. In particular, for the ammonia, methylamine, dimethylamine, trimethylamine series, the errors in the theory were found to be correlated with the strength of the  $\text{NH}\cdots\text{O}$  hydrogen bonds in separate, gas phase calculations carried out on the  $\text{H}_2\text{O}$ -solute system. A “solvent accessible” correction term has been proposed to account for these effects. Denoting  $\psi^g$  the gas phase wave function, and with  $A$  given in  $\text{\AA}^2$ , the separate contributions are given by

$$\begin{aligned} \Delta G^{es} &= \langle \psi^s | H^0 | \psi^s \rangle - \langle \psi^g | H^0 | \psi^g \rangle + \frac{1}{2} [\langle \psi^s | H' | \psi^s \rangle + H''], \\ \Delta G^{np} &= 1.09 + 0.005 A, \end{aligned} \quad (2.3)$$

$$\Delta G^{sa} = \sum_i a_i e^{-br_i^2}.$$

Here  $b$  is an empirically adjusted parameter with value  $10 \text{\AA}^2$  and  $r_i$  is the distance between the point on the molecular surface closest to the atom associated with functional group  $i$  and its dielectric sphere, as defined by the assigned radius. The parameter  $a_i$  is the solvent accessible correction factor for functional group  $i$ . The hydrogen bonding correction term for an entire molecule is assumed to be the sum of the solvent accessible correction terms for the different functional groups on the molecule. Also, for the choice of parameters used in Eq. (2.3),  $\Delta G^{np}$  is given in kcal/mol.

## III. 3D FINITE ELEMENT FORMULATION OF THE DIELECTRIC CONTINUUM PROBLEM

A number of different numerical methods have been proposed over the years for solving the Poisson–Boltzmann equation in molecular geometries. Three-dimensional methods have been based on the use of cubic grids and finite difference (FD) representations of the partial differential equation.<sup>16</sup> Two-dimensional boundary element methods (BEM) on the other hand use a discretization of the molecular surface and have been combined with three-dimensional finite elements to include the effects of the ionic atmosphere in calculations.<sup>17</sup> By using a highly nonuniform adaptive finite element mesh, we have retained the low connectivity property characteristic of three-dimensional methods which leads to a sparse matrix representation of the discretized

problem. The adaptivity allows us to concentrate the grid point distribution near the internal boundary in the problem, so that the computational cost of the method scales with the complexity of the molecular surface rather than with the molecular size.

The grid point distribution is generated using a combination of Lebedev quadrature grids centered on the atoms and an exact construction of the reentrant sections of the molecular surface as defined in the Richards<sup>18</sup> construction. An adaptive finite element mesh is then obtained using a constrained Delaunay triangulation/marching-front algorithm developed specifically for the problem of rapidly calculating the induced surface charge distribution. The reader is referred to the fluid dynamics and numerical methods literature for a detailed discussion of the general mesh generation problem, for example see the work of Baker.<sup>19</sup>

We turn to the discretization problem. Consider the linear Poisson–Boltzmann case

$$\nabla \cdot (\epsilon \nabla \phi) = -\frac{4\pi\rho}{kT} + \epsilon\kappa^2\phi, \quad (3.1)$$

where  $\phi(\mathbf{r})$  is given in units of  $[kT]$ , and  $\kappa^2 = (8\pi e^2 I) / (\epsilon kT)$  is the inverse Debye length squared. We will focus on the linear form of the equation hereafter, as the finite element method is directly applicable to it. We wish to discretize the problem using a finite computational space. Given a complete set of basis functions  $\{\psi_i\}$   $i = 1 \cdots N$ , one can expand the electrostatic potential

$$\phi(\mathbf{r}) = \sum_{i=1}^N c_i \psi_i(\mathbf{r}). \quad (3.2)$$

We choose the set of functions  $\{\psi_i\}$  to be a finite element basis. This means each  $\psi_i$  is nonzero only in some finite region around  $\mathbf{r}_i$  and satisfies the following conditions:

$$\begin{aligned} \psi_i(r_j) &= 1 & \text{if } i=j, \\ \psi_i(r_j) &= 0 & \text{if } i \neq j. \end{aligned} \quad (3.3)$$

For practical purposes, we also choose the basis functions to be linear functions of the Cartesian coordinates. This choice is not necessary in principle.

In Eq. (3.2), the coefficients  $c_i$  are the values of the potential at the grid vertices. The expansion is substituted into Eq. (3.1). Multiplying by  $\psi_j$ , integrating over all space, using Green's identity and the divergence theorem as well as the vanishing condition for the potential at infinity, we obtain a set of linear equations

$$\sum_i^N A_{ij} c_i = b_j, \quad (3.4)$$

where

$$\begin{aligned} A_{ij} &= -\int_V \nabla(\epsilon\psi_i) \cdot \nabla\psi_j d^3\mathbf{r} - \int_V \epsilon\kappa^2\psi_i\psi_j d^3\mathbf{r}, \\ b_j &= -\int_V \frac{4\pi\rho}{kT} \psi_j d^3\mathbf{r}. \end{aligned} \quad (3.5)$$

The finite size of the computational space is handled through the use of the Debye–Hückel potential at the outer boundary of the grid. The system of equations is then solved iteratively using a Jacobi conjugate gradient iteration scheme.

Within the dielectric continuum model, the polarization contribution to the free energy of solvation  $\Delta G^{\text{pol}}$  is simply the cost of inserting an object of dielectric constant  $\epsilon$  in an external field  $\mathbf{E}_0$ . This is written:

$$\Delta G^{\text{pol}} = -\frac{1}{2} \int_V (\mathbf{P} \cdot \mathbf{E}_0) d^3\mathbf{r}, \quad (3.6)$$

where  $\mathbf{P} = \chi_e \mathbf{E}$  is the polarization field in the solvent continuum and  $\mathbf{E}_0$  is the electric field produced by the fixed charge distribution of the solute. Using the divergence theorem, and the internal boundary conditions:

$$\epsilon_0 \mathbf{E}^i \cdot \hat{\mathbf{n}} = \epsilon \mathbf{E}^o \cdot \hat{\mathbf{n}}, \quad (3.7)$$

where  $\hat{\mathbf{n}}$  is the outwards directed normal to the molecular surface,  $\mathbf{E}^i$  and  $\mathbf{E}^o$  are the interior and exterior electric fields at the surface,  $\Delta G^{\text{pol}}$  can be written:

$$\begin{aligned} \Delta G^{\text{pol}} &= -\frac{1}{8\pi} \int_S (\mathbf{n} \cdot \mathbf{E}^i - \mathbf{n} \cdot \mathbf{E}^o) \phi_0 d^3\mathbf{r} \\ &\quad + \frac{1}{2} \int_V (\nabla \cdot \mathbf{P}) \phi_0 d^3\mathbf{r}. \end{aligned} \quad (3.8)$$

Also, we have used  $\mathbf{E}_0 = -\nabla\phi_0$  in the last equation. The expression for  $\Delta G^{\text{pol}}$  can be rewritten in terms of a surface polarization charge density  $\sigma$  and the density of free charges  $\rho_{\text{ion}}$ . The charges  $\sigma$  are distributed on the molecular surface while the density  $\rho_{\text{ion}}$  is nonzero only beyond the ion accessible surface in the solvent and is linear in the potential  $\phi$  for the case of the linear PB equation. We obtain

$$\Delta G^{\text{pol}} = \frac{1}{2} \int_S \sigma \phi_0 d^3\mathbf{r} + \frac{1}{2} \int_V \rho_{\text{ion}} \phi_0 d^3\mathbf{r}. \quad (3.9)$$

Evaluation of the first term in Eq. (3.9) requires calculating the surface polarization charge:

$$\sigma(\mathbf{r}) = \frac{1}{4\pi} \mathbf{n} \cdot (\nabla\phi^i - \nabla\phi^o). \quad (3.10)$$

The method has been implemented in our Poisson–Boltzmann finite element solver (PBF). As will be detailed in a separate publication, the results obtained for electrostatic contributions to solvation free energies are in good agreement with those from other programs such as DELPHI.<sup>16</sup> Also, the robustness of the mesh generation algorithm, a frequent obstacle in the large-scale use of unstructured three-dimensional meshes, has been tested on a database of 550 different compounds and succeeded in producing usable triangulations for each one.

#### IV. POLARIZATION ENERGY GRADIENT CALCULATION

##### A. Choice of method

Recently, Gilson *et al.*<sup>20</sup> have reported preliminary results for an analytical gradient methodology for classical PB applications. In their calculations, there are some problems attaining high precision gradient components although the computational cost of the method is quite low. In the present chapter, we describe a very different formalism which can produce results of high accuracy at a CPU cost comparable to that of a single point energy calculation. The gradient is obtained from exact linear algebra relations by using a Green's function formulation due to Handy and Schaefer<sup>21</sup> which they implemented in the context of electronic structure theory. The methodology developed for calculating the gradient is outlined below. Comparisons with energy finite differences are performed for the stand-alone dielectric continuum problem in order to assess both the accuracy and efficiency of the method. Additional comparisons between our implementation of Gilson's formalism and the method described here have also been carried out and will be detailed in another publication. It should be noted that we have found it possible to obtain gradients extremely rapidly using Gilson's formalism, though the method appears to be less accurate than the algorithm described below.

##### B. Mesh operator derivative method (MOD)

As has been outlined above, in the linear PB equation framework the electrostatic contribution to the solvation free energy is given by

$$\Delta G^{\text{pol}} = \frac{1}{2} \int_S \sigma \phi d^3 \mathbf{r} + \frac{1}{2} \int_V \rho_{\text{ion}} \phi d^3 \mathbf{r}. \quad (4.1)$$

Where  $\rho_{\text{ion}}$ , the density of free ions is linear in  $\phi$ , and is nonzero only beyond the ion-accessible surface. The discretized form of this equation arising from the finite element representation of the problem we have developed can be written as

$$\Delta G^{\text{pol}} = \mathbf{z} \cdot \tilde{\phi} + \mathbf{y} \cdot \tilde{\phi}^2, \quad (4.2)$$

where  $\mathbf{z}$ ,  $\mathbf{y}$  are vectors of integration weights depending on the coordinates of grid vertices and  $\tilde{\phi}$ ,  $\tilde{\phi}^2$  are vectors containing the first and second powers of the grid potential values. Let  $q$  be any nuclear Cartesian coordinate. Differentiating Eq. (4.2) with respect to  $q$  we have

$$\frac{\partial \Delta G^{\text{pol}}}{\partial q} = \frac{\partial \mathbf{z}}{\partial q} \tilde{\phi} + \frac{\partial \mathbf{y}}{\partial q} \tilde{\phi}^2 + \mathbf{z} \frac{\partial \tilde{\phi}}{\partial q} + 2 \tilde{\phi} (\mathbf{y}^T \cdot I) \frac{\partial \tilde{\phi}}{\partial q}, \quad (4.3)$$

where  $I$  is the identity matrix. Similarly, differentiating the system of linear equations obtained from the discretization of the Dirichlet problem (3.4) we obtain

$$\frac{\partial A}{\partial q} \tilde{\phi} + A \frac{\partial \tilde{\phi}}{\partial q} = \frac{\partial \mathbf{b}}{\partial q}. \quad (4.4)$$

Solving Eq. (4.4) for  $\partial \tilde{\phi} / \partial q$  then substituting in Eq. (4.3) we have

$$\begin{aligned} \frac{\partial \Delta G^{\text{pol}}}{\partial q} &= \frac{\partial \mathbf{z}}{\partial q} \tilde{\phi} + \frac{\partial \mathbf{y}}{\partial q} \tilde{\phi}^2 + [\mathbf{z} + 2 \tilde{\phi} (\mathbf{y}^T \cdot I)] A^{-1} \\ &\quad \times \left( \frac{\partial \mathbf{b}}{\partial q} - \frac{\partial A}{\partial q} \tilde{\phi} \right). \end{aligned} \quad (4.5)$$

In principle, constructing the inverse operator  $A^{-1}$  would require a direct solution of the system of Eq. (3.4). This would be prohibitively expensive as even for small molecules, the grids are composed of several thousand vertices. Instead, using the symmetry of the operator  $A$  one can iteratively solve the linear system:

$$A \cdot \mathbf{g} = \mathbf{z} + 2 \tilde{\phi} (\mathbf{y}^T \cdot I). \quad (4.6)$$

Equation (4.18) then becomes

$$\frac{\partial \Delta G^{\text{pol}}}{\partial q} = \frac{\partial \mathbf{z}}{\partial q} \tilde{\phi} + \frac{\partial \mathbf{y}}{\partial q} \tilde{\phi}^2 + \mathbf{g} \left( \frac{\partial \mathbf{b}}{\partial q} - \frac{\partial A}{\partial q} \tilde{\phi} \right). \quad (4.7)$$

The adaptive grid structure now makes it possible to numerically evaluate the various derivative terms appearing in Eq. (4.7). The coordinates of every vertex in the grid are functions of the nuclear coordinates of one, two, or three atoms at most. It is therefore possible, to compute the derivatives by finite differencing the terms  $\mathbf{z}$ ,  $\mathbf{y}$ ,  $\mathbf{b}$ , and  $A$  with respect to small atomic displacements. Since only a subset of the components of these terms need to be updated for a given atomic displacement, this approach is relatively inexpensive. In fact, most of the time is spent in compressing the information needed to update the operator  $A$  in order to minimize the storage requirements of the method (only the nonzero elements of  $A$  are stored). An alternate approach, involving a single compression loop for the mesh operator, can be obtained by decomposing the derivatives as follows:

$$\frac{\partial w_i}{\partial q} = \sum_k \frac{\partial w_i}{\partial p_k} \frac{\partial p_k}{\partial q}, \quad (4.8)$$

$$\frac{\partial a_{ij}}{\partial q} = \sum_k \frac{\partial a_{ij}}{\partial p_k} \frac{\partial p_k}{\partial q}, \quad (4.9)$$

where  $w_i$  is any component of  $\mathbf{z}$ ,  $\mathbf{y}$ ,  $\mathbf{b}$ ,  $a_{ij}$  is any element of  $A$  and  $p_k$  is a Cartesian coordinate of any vertex in the grid. With this approach, the derivatives of the matrix elements with respect to vertex coordinates  $\partial a_{ij} / \partial p_k$  can be preprocessed so that only the derivatives  $\partial p_k / \partial q$  need to be recomputed for every nuclear displacement. This eliminates the data compression cost at each step but requires the calculation of a large number of terms as a preprocessing stage. As can be seen from Eq. (3.5), each matrix element  $a_{ij}$  depends on 10 inner products of the linear basis functions. Since the 12 possible derivatives of each inner product with respect to the Cartesian coordinates  $\{p_k\}$  of the different vertices in a given tetrahedron must be computed, a total of 120 different expressions for each matrix element must be coded. These are evaluated during the preprocessing step.

TABLE I. Gradient components in units of  $[kT]/\text{\AA}$  at 298 K for water. Results from the MOD method are shown at various grid densities determined by the parameter  $l_{sf}$ . The sum of gradient components in each direction is listed in the last three columns. The results from energy finite differences are given in the last two rows. The atomic displacement  $\Delta q$  is given in  $\text{\AA}$ .

$l_{sf}$	$O_x$	$O_y$	$O_z$	$H1_x$	$H1_y$	$H1_z$	$H2_x$	$H2_y$	$H2_z$	$\Sigma_x$	$\Sigma_y$	$\Sigma_z$
8	-0.02	0.20	35.45	0.27	-6.42	-17.72	-0.25	6.22	-17.73	0.00	0.00	0.00
11	0.78	-0.19	35.73	-0.47	-5.88	-17.84	-0.31	6.07	-17.89	0.00	0.00	0.00
14	0.06	0.15	37.63	-0.02	-3.99	-18.80	-0.04	3.84	-18.83	0.00	0.00	0.00
16	0.25	-0.18	35.01	-0.16	-5.17	-17.51	-0.09	5.35	-17.50	0.00	0.00	0.00
19	-0.02	0.01	35.19	-0.02	-4.72	-17.60	0.04	4.71	-17.59	0.00	0.00	0.00
$\Delta q$												
0.05	-0.02	0.06	35.15	0.01	-4.85	-17.87	0.02	4.86	-18.02	0.01	0.07	-0.74
0.10	-0.01	0.08	35.03	0.01	-4.29	-17.62	0.01	4.17	-17.57	0.01	-0.04	-0.16

### C. Numerical results for the classical polarization energy gradient

Some numerical experiments have been performed to assess the accuracy of the gradient methodology we have developed for the isolated dielectric continuum problem. Results of accuracy tests for the full quantum mechanical solution phase energy gradient will be discussed below. We compare gradient components for water obtained by energy finite differences, and the MOD method. The results obtained are listed in Table I. The parameter  $l_{sf}$  refers to the Lebedev grid index controlling the resolution used in the Richards surface and 3D molecular grid generation. Larger values of  $l_{sf}$  indicate a higher grid point density.

The results indicate that the gradient components obtained with the MOD method are in good agreement with those computed by energy finite differences. Furthermore, a necessary property of any energy gradient method is that the net force acting on the system coming from accumulation of error, and which must be zero formally, be as small as possible. Though it is possible to remove any net translation or rotation of the system by rescaling the forces at the end of the calculation, this tends to introduce some noise in the minimization procedure so that it is desirable to reduce this effect. In this respect, it is apparent that the MOD calculations present a very attractive feature. With the grids employed above, the cancellation of error systematically results in net gradient components inferior to  $5 \cdot 10^{-3} [kT]/\text{\AA}$ . For the simple case considered here, the CPU time necessary to obtain gradients from the MOD method was 2.7 s on an IBM370 workstation for the  $l_{sf}=19$  case, about the same time required for the single point energy finite element calculation itself. In the context of SCRF methods, where the continuum electrostatic part of the calculation represents less than 10% of the CPU time of the complete calculation, this performance is quite acceptable. As we have remarked above, because of its greater efficiency, our implementation of Gilson's formalism should nevertheless be the algorithm of choice for use in applications involving molecular mechanics force fields where the accuracy requirements are not as stringent as in the context of SCRF applications.

## V. GEOMETRY OPTIMIZATION IN THE CONTEXT OF SCRF METHODS

### A. Solution phase energy gradient

The methodology for performing geometry optimization calculations using electronic structure methods has now been in place for many years and the algorithm used for the quantum mechanical part of our calculations has been described in a paper by Won *et al.*<sup>22</sup> In principle, similar algorithms can be used to carry out simulations in solution provided a gradient of the interaction between the molecule and solvent reaction field with respect to the nuclear coordinates is available. Here, the solvent reaction field is taken to be the Coulomb potential of the induced surface polarization charge distribution  $\sigma$ . The molecular configuration sought is the one which minimizes the total solution phase energy  $E_s^T$  of the system. The latter is the sum of the quantum mechanical solution phase energy  $E_s^{\text{QM}}$  and the nonpolar hydrophobic term  $\Delta G^{\text{np}}$  (for the moment we will assume no hydrogen bonding term is used in the calculation of the solvation free energy):

$$E_s^T = E_s^{\text{QM}} + \Delta G^{\text{np}}. \quad (5.1)$$

The geometry of the molecule is updated using the gradient of Eq. (5.1), with respect to the nuclear coordinates and the procedure is then iterated until the change in solution phase energies between successive configurations lies below some predetermined value.

The method used to combine the classical and quantum mechanical gradients, to obtain derivatives of the quantum mechanical solution phase energy functional, can now be summarized in a few steps. Differentiating Eq. (5.1) with respect to a nuclear Cartesian coordinate  $q$  we obtain

$$\frac{\partial E_s^T}{\partial q} = \frac{\partial}{\partial q} (\langle \psi^s | H^0 | \psi^s \rangle) + \frac{1}{2} \frac{\partial}{\partial q} (\langle \psi^s | H' | \psi^s \rangle + H'') + \frac{\partial \Delta G^{\text{np}}}{\partial q}. \quad (5.2)$$

The first term involves differentiation of the kinetic energy and solute-solute interaction part of the Hamiltonian, and will not be discussed further in this article. A detailed de-

scription of this term for HF level calculations is given by Won *et al.*<sup>22</sup> The methodology for GVB and MP2 level gradients will be described in upcoming articles. The second term is just the derivative of the polarization contribution to the free energy of solvation. As written above, it is different from Eq. (4.3) since the reaction field here is assumed to correspond to the solvent response to a continuous charge distribution. The source term used in our continuum model calculations, however, corresponds to a point charge representation obtained by ESP fitting of molecular charge distribution.

Provided the solution to the classical electrostatics problem for the discrete sources is a good approximation to the solution for the continuous ones in the neighborhood of the dielectric interface, the reaction field obtained from our finite element calculations can be used to approximate  $H' + H''$ . This assumption seems reasonable as it has been verified that near the dielectric interface the Coulomb field from the ESP fitted charges differs from the Coulomb field of the molecular charge distribution by less than 5%.

The quantum mechanical expression for the gradient with respect to an atomic coordinate given by Eq. (5.2), can be viewed as the sum of two terms. The first arises from the interaction between the reaction field due to the induced charge and the quantum mechanical molecular charge distribution at a *fixed* molecular surface configuration. We expect this term, which includes information from the complete molecular charge distribution, to be a better representation of the force due to interactions between the molecule and the induced surface charge than the corresponding term in the classical force expression. The second term accounts for the dependence of the molecular surface (which is updated during the minimization procedure) on the atomic coordinates. To avoid double counting, if the full expression (5.2) is used in the gradient calculation, the corresponding classical term must be subtracted from the total gradient. The remaining term, which accounts for the pressure exerted by the solvent dielectric continuum on the Richards surface of the molecule, is just the second term mentioned above. More precisely, we replace the classical reaction field Coulomb term given by

$$\frac{\partial \Delta G_r^{\text{pol}}}{\partial q} = \frac{1}{2} \frac{\partial}{\partial q} \left( \sum_i \int_S \frac{\sigma(\mathbf{r}')}{\|\mathbf{r}_i - \mathbf{r}'\|} Q_i d^2 \mathbf{r}' \right)_S, \quad (5.3)$$

with its quantum mechanical counterpart:

$$\begin{aligned} \frac{\partial \Delta G_r^{\text{pol,QM}}}{\partial q} = & \frac{1}{2} \frac{\partial}{\partial q} \left( \int_S \psi^*(\mathbf{r}') \frac{\sigma(\mathbf{r}')}{\|\mathbf{r} - \mathbf{r}'\|} \psi(\mathbf{r}') d^2 \mathbf{r}' \right. \\ & \left. + \sum_i \int_S \frac{\sigma(\mathbf{r}')}{\|\mathbf{r}_i - \mathbf{r}'\|} Z_i d^2 \mathbf{r}' \right)_S. \end{aligned} \quad (5.4)$$

The subscript  $S$  indicates that the derivatives are to be taken at a *fixed* configuration of the Richards surface. This contrasts with the meaning of the derivative in Eq. (5.2), where the differentiation takes into account the dependence of the molecular surface on the atomic coordinates. In the latter equations,  $Q_i$  designates the ESP fitted charge on atom  $i$ ,  $Z_i$

the nuclear charge. Finally, in the case of zero ionic concentration, the second term on the right-hand side in Eq. (5.2), which will be used in geometry optimizations, is explicitly given by

$$\begin{aligned} \frac{\partial \Delta G^{\text{pol,QM}}}{\partial q} = & \frac{\partial \Delta G_r^{\text{pol,QM}}}{\partial q} - \frac{\partial \Delta G_r^{\text{pol}}}{\partial q} + \frac{\partial \mathbf{z}}{\partial q} \phi \\ & + \mathbf{g} \left( \frac{\partial \mathbf{b}}{\partial q} - \frac{\partial A}{\partial q} \phi \right). \end{aligned} \quad (5.5)$$

We use the SCRF method, developed by Tannor and co-workers,<sup>5</sup> as implemented in the PS-GVB system of programs<sup>23</sup> for the quantum mechanical part of the calculation and the optimization procedure.

## B. Solution phase geometry optimization: HF level test calculations

Having established the validity of our gradient for the classical electrostatic solvation energy function, we now turn to the verification of the combined SCRF-PBF method. The numerical reliability of Eq. (5.5) needs to be assessed before considering applications of the method to molecular modeling problems. In all SCRF calculations discussed in the current and following sections, the solvent dielectric continuum is assigned a dielectric constant of 78.3 and the probe radius  $r_p$  is set to 1.4 Å to represent an aqueous environment. The quantum mechanical parts of the calculation were done at HF/6-31G\*\* level. A converged geometry is defined by maximum values of less than  $2.25 \times 10^{-3}$ ,  $1.5 \times 10^{-3}$  [Hr/bohr] for the gradient components and r.m.s of gradient elements, maximum values of less than  $5.0 \times 10^{-2}$  [bohr] or [rad] for both the maximum and r.m.s of nuclear displacement elements and total free energy differences of less than  $2.5 \times 10^{-4}$  [Hr] between previous and current geometry optimization iterations.

It is well known that Hartree–Fock level calculations on salts fail to predict dissociation in solution. Instead, the minimum ground-state energy for the solvated system is found at a finite bond length. A line search along the bond-length coordinate, consisting of a series of single point energy calculations, reveals that the equilibrium bond length for Li–F lies between 1.665 and 1.670 Å corresponding to the minimum of the solution phase energy (Fig. 1). Accordingly, at that bond length the system has moved away from the gas phase equilibrium (Fig. 2). A first test of the energy gradient function is done by comparing the energy derivatives obtained by finite differencing the solution phase energy at the gas phase equilibrium point and comparing with the predicted value. The results for central differencing along the bond direction are shown in Table II.

In test (b), the solvation calculations were performed using a mesh approximately eight times as dense as in cases (a). In both cases, there is good agreement between the gradient and the finite difference calculations for the smaller value of the nuclear displacement. Also, using our SCRF-PBF energy gradient to drive an internal coordinate geometry optimization calculation, we find an equilibrium bond

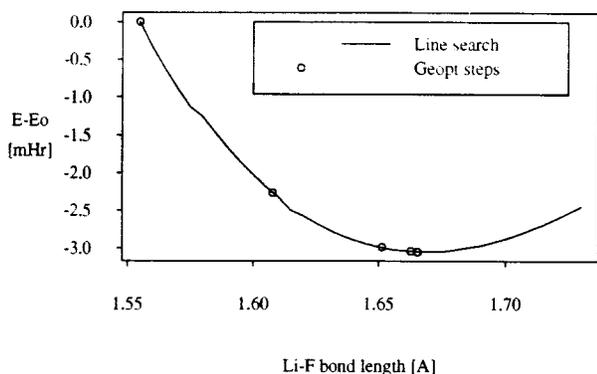


FIG. 1. Location of geometry optimization steps along the Li-F bond during the solution phase energy minimization process. The solid line, obtained from a series of single point energy calculations, indicates the dependence of the change in solution phase energy on the bond length.  $E-E_0$  is the change in solution phase energy from the value at the gas phase equilibrium geometry.

length of 1.6653 Å for the system. The implication is that the use of Eq. (5.5) results is a valid approximation in the context of assembling a gradient of the total SCR-F-PBF energy at the HF level, and can be used to determine local energy and geometry changes between gas phase and solution phase equilibrium configurations.

### C. Solvation effects on the molecular structure of organic compounds

We now present the results obtained in a series of five solution phase geometry optimizations of small organic compounds. The molecules studied are shown in Fig. 3. These are in order, acetamide, *N*-methylacetamide, acrolein, mesityl oxide, and one example of a merocyanine dye. Two resonance structures are shown for each compound. The molecule in the gas phase can be represented as a hybrid of these two resonance structures resembling more the one on the left. In the solvated molecule the equilibrium is shifted towards the right, therefore the contribution of the zwitterionic structure is increased. In this section, we examine only the magnitude of the structural and energetic changes introduced

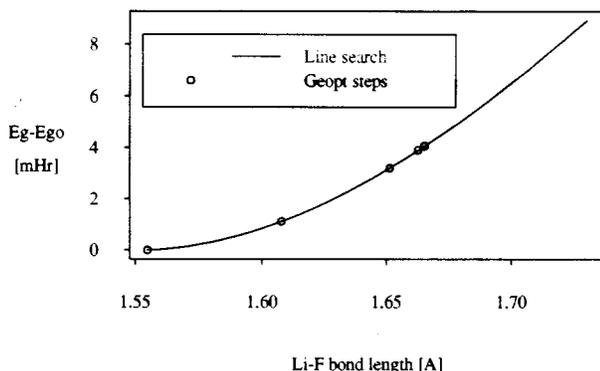


FIG. 2. Notation as in Fig. 1. Here gas phase energy change is shown at each geometry optimization step. The solid line is the result of a line search.

in the solution phase optimization process. The issue of comparison with other methods or experiments will be briefly addressed in the following section and is the topic of ongoing work. All the changes discussed below are measured between the solvated gas phase optimized geometries and the solution phase optimized geometries.

Tables III and IV list the structural changes observed between the gas phase and solution phase minima of the SCR-F-PBF energies at the HF/6-31G\*\* level. With regards to bond length, or “stretch” degrees of freedom, it is apparent that, where nontrivial changes are observed, the solvation process results in a lengthening of double bonds and a shortening of single bonds suggesting that the picture described by the interconversion between the structures shown in Fig. 3 is at least qualitatively correct. More precisely, for compounds (A)–(D) we observe the largest bond-length changes for C=O and C–N bonds in a direction which suggests evolution to C–O and C=N like structures. The changes in the ESP fitted partial charges listed in Table V are consistent with this charge separation mechanism. For compounds (A)–(D), no other significant structural modifications are observed as the largest angle changes take place in “bend” degrees of freedom and do not exceed 1.2 deg (see Tables III and IV). Correspondingly, only small changes in the free energy of solvation are observed. These are listed in Table VI.

The case of the merocyanine dye (E) appears more interesting in the light of the 6.4 kcal/mol change in solvation energy observed. The large energy change can be understood by noting first, that the structural changes for (E) are about 2–4 times larger than those observed for (A)–(D) (see Table IV). Also, the coupled changes in bond lengths observed along the longest axis of the molecule, appear to induce a significant charge separation effect over a distance of several bond lengths. This is manifest in the changes in the ESP fitted partial charges located on the O, C<sub>7</sub>, and C<sub>14</sub> atoms, as shown in Table V. The effect can be visualized, by examining identical contour lines of the electrostatic potential of the molecule within the dielectric continuum model for the different conformations and ESP fitted charges. These are displayed in Figs. 4 and 5 for the initial and final configurations, respectively. The line of zero electrostatic potential shifts away from the N side of the structure in the minimization process and appears to relocate in the region of the C<sub>7</sub> car-

TABLE II. Comparison of total solution phase free energy gradient obtained using the MOD method  $\partial E_{\text{sol}}/\partial q$  with gradient from energy finite differences  $(\Delta E_{\text{sol}})/(2\Delta q)$ . The total atomic displacement used is  $2\Delta q$ . Results are shown for two different grid resolutions differing by a factor of 8 in grid point density, with (b) the densest grid.

Mesh type	$\Delta q$ [Å]	$\frac{\Delta E_{\text{sol}}}{2\Delta q}$ [Hr/Å]	$\frac{\partial E_{\text{sol}}}{\partial q}$ [Hr/Å]
a	0.05	3.481e-02	3.209e-02
a	0.025	3.722e-02	3.209e-02
b	0.05	3.400e-02	3.228e-02
b	0.025	3.386e-02	3.228e-02

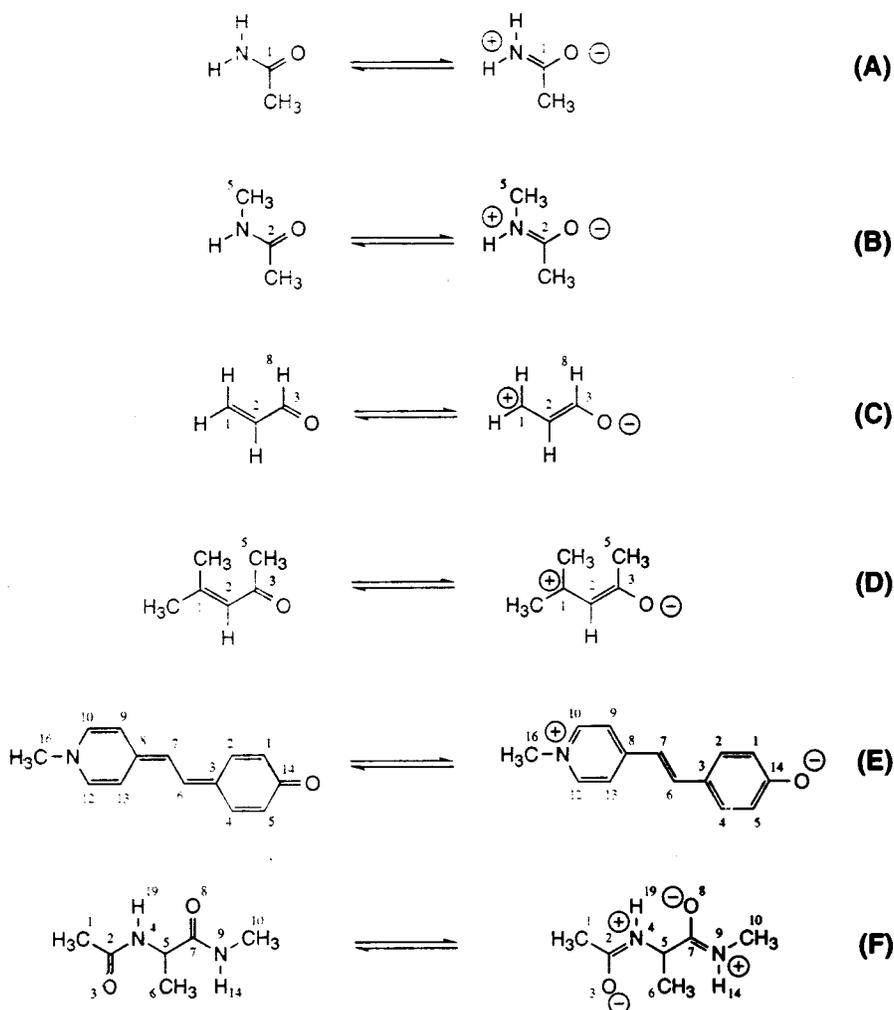


FIG. 3. Resonance structures of compounds used in solution phase geometry optimization tests. (A) acetamide, (B) *N*-methylacetamide, (C) acrolein, (D) mesityl oxide, (E) merocyanine, (F) 1-(acetylamino)-*N*-methylpropanamide (AD).

bon, reflecting the large change in electrostatic charge in this region. The smaller scale structure in the potential also appears to be attenuated in solution and the tighter spacing of the contour lines on either ends indicate that an accumulation of negative charge has taken place in the O region and of positive charge in the N region. In this respect, the changes are qualitatively similar to those observed for the smaller compounds, however, the alternation of single and double bonds along the entire structure allows for greater charge delocalization and response to the solvent environment. It is likely that the conclusions drawn from this example are relevant to the general problem of modeling solvated compounds containing aromatic rings, or more generally several conjugated double bonds, regardless of their apparent flexibility (rotational freedom about torsion angles).

## VI. APPLICATION TO THE STUDY OF THE ALANINE DIPEPTIDE

The alanine dipeptide (compound F in Fig. 3) or 1-(acetylamino)-*N*-methylpropanamide, has been the subject

of a number of studies using both molecular mechanics force fields and *ab initio* methods. Several references to these numerical experiments are provided by Gould and co-workers.<sup>1</sup> Because of its basic structural similarity to larger peptides, quantum mechanical calculations performed on this system should prove useful in assessing the validity of classical force field methods in the context of applications to proteins. The increase in efficiency of *ab initio* methods observed in the last few years has already allowed systematic studies of the potential energy surface of this system and related ones to be carried out. In particular, Head-Gordon *et al.*<sup>24</sup> have reported high resolution studies of the Ramachandran map for glycine and alanine peptide analogs in the gas phase (the compound used in the Head-Gordon study differs somewhat from (F) in that the terminal methyl groups C<sub>1</sub>, C<sub>10</sub> are replaced by hydrogens.) Until recently however, only gas phase *ab initio* studies were available and a more rigorous treatment of solvation effects has been lacking. Gould *et al.*<sup>1</sup> have carried out solvated *ab initio* calculations at the HF/6-31G\*\* level, using both an SCRF method developed

TABLE III. Structural changes for molecules (A)–(D). Changes in bond lengths are shown for all bonds not involving hydrogens. Bend angle changes are shown when they are greater than 1°. No torsion angle changes greater than 1° were observed for (A)–(D).

Molecule	Coordinate $q$	$\Delta q$ [Å],[deg.]
Acetamide	N–C <sub>1</sub>	–0.018 35
	C <sub>1</sub> –C <sub>4</sub>	–0.004 79
	C <sub>1</sub> –O	0.018 91
	C <sub>4</sub> –C <sub>1</sub> –N	1.035 67
	H <sub>6</sub> –N–C <sub>1</sub>	1.084 48
<i>N</i> -methylacetamide	C <sub>1</sub> –C <sub>2</sub>	–0.004 65
	C <sub>2</sub> –O	0.019 67
	N–C <sub>5</sub>	0.007 71
	C <sub>2</sub> –N	–0.019 95
	H <sub>9</sub> –N–C <sub>2</sub>	1.158 69
	N–C <sub>2</sub> –C <sub>1</sub>	1.063 92
Acrolein	C <sub>1</sub> –C <sub>2</sub>	0.003 45
	C <sub>2</sub> –C <sub>3</sub>	–0.008 94
	C <sub>3</sub> –O	0.011 74
Mesityl oxide	C <sub>1</sub> –C <sub>2</sub>	0.004 10
	C <sub>1</sub> –C <sub>7</sub>	–0.000 14
	C <sub>3</sub> –C <sub>5</sub>	–0.004 83
	C <sub>1</sub> –C <sub>6</sub>	–0.000 51
	C <sub>2</sub> –C <sub>3</sub>	–0.007 67
	C <sub>3</sub> –O	0.010 10

TABLE IV. Structural changes for molecule (E). Changes in stretches are shown for all bonds not involving hydrogens. Bend angle changes are shown when they are greater than 1° and do not involve hydrogens. No torsion angle changes greater than 1° were observed.

Molecule	Coordinate $q$	$\Delta q$ [Å],[deg.]
Merocyanine	C <sub>1</sub> –C <sub>2</sub>	0.029 19
	C <sub>3</sub> –C <sub>6</sub>	0.077 73
	C <sub>8</sub> –C <sub>13</sub>	–0.042 91
	N–C <sub>16</sub>	0.021 78
	C <sub>14</sub> –O	0.048 63
	C <sub>1</sub> –C <sub>14</sub>	–0.038 42
	C <sub>2</sub> –C <sub>3</sub>	–0.042 66
	C <sub>3</sub> –C <sub>4</sub>	–0.046 33
	C <sub>4</sub> –C <sub>5</sub>	0.033 54
	C <sub>5</sub> –C <sub>14</sub>	–0.040 60
	C <sub>6</sub> –C <sub>7</sub>	–0.078 59
	C <sub>7</sub> –C <sub>8</sub>	0.079 18
	C <sub>8</sub> –C <sub>9</sub>	–0.043 75
	C <sub>9</sub> –C <sub>10</sub>	0.023 58
	C <sub>10</sub> –N	–0.023 09
	N–C <sub>12</sub>	–0.026 86
	C <sub>12</sub> –C <sub>13</sub>	0.024 79
	C <sub>13</sub> –C <sub>8</sub> –C <sub>7</sub>	–1.614 24
	C <sub>10</sub> –C <sub>9</sub> –C <sub>8</sub>	–1.102 82
	C <sub>12</sub> –N–C <sub>10</sub>	1.894 77
	C <sub>12</sub> –C <sub>13</sub> –C <sub>8</sub>	–1.265 20
	C <sub>9</sub> –C <sub>8</sub> –C <sub>7</sub>	–1.353 07
	C <sub>13</sub> –C <sub>8</sub> –C <sub>9</sub>	2.967 31
	N–C <sub>10</sub> –C <sub>9</sub>	–1.340 16
	C <sub>16</sub> –N–C <sub>10</sub>	–1.154 23
	C <sub>13</sub> –C <sub>12</sub> –N	–1.154 02

TABLE V. Changes in ESP fitted charges for molecules (A)–(E). For hydrogens only the total change in charge is shown.

Molecule	Atom	$\Delta q$ [ $e$ ]
Acetamide	N	0.027
	O	–0.019
	C	–0.007
<i>N</i> -methylacetamide	{H}	–0.001
	C <sub>1</sub>	0.003
	C <sub>2</sub>	–0.007
	O	–0.023
	N	0.024
Acrolein	C <sub>5</sub>	–0.021
	{H}	0.024
	C <sub>1</sub>	0.006
	C <sub>2</sub>	–0.001
	C <sub>3</sub>	–0.007
Mesityl oxide	O	–0.008
	{H}	0.010
	C <sub>1</sub>	0.003
	C <sub>2</sub>	–0.001
	C <sub>3</sub>	–0.002
Merocyanine	O	–0.007
	C <sub>5</sub>	0.001
	C <sub>6</sub>	–0.001
	C <sub>7</sub>	–0.002
	{H}	0.009
	C <sub>1</sub>	0.008
	C <sub>2</sub>	–0.062
	C <sub>3</sub>	0.006
	C <sub>4</sub>	–0.028
	C <sub>5</sub>	–0.013
C <sub>6</sub>	–0.047	
C <sub>7</sub>	0.125	
C <sub>8</sub>	–0.047	
C <sub>9</sub>	0.039	
C <sub>10</sub>	0.057	
N	0.019	
C <sub>12</sub>	0.052	
C <sub>13</sub>	0.055	
C <sub>14</sub>	–0.112	
O	–0.111	
C <sub>16</sub>	0.006	
{H}	0.053	

by Rivail *et al.*<sup>7–9</sup> (which will be referred to hereafter as SCRF-R) and the polarizable continuum model (PCM) developed by Tomasi and co-workers<sup>25</sup> on these two dipeptides. In the SCRF-R method, with which the reported solvated geometry optimizations were carried out, the solute cavity is represented simply as an elliptical cavity and the solute charge distribution as a finite multipole expansion of the molecular charge distribution. Our calculations outline the effects of using a more realistic solvent model on the location and heights of the minima in the potential energy surface of the alanine dipeptide.

The conformations discussed refer to different values of the ( $\Phi, \Psi$ ) internal angles for the structure. These correspond to different rotational configurations of the backbone of the structure and correspond to the (H<sub>19</sub>N<sub>4</sub>C<sub>5</sub>C<sub>7</sub>), (N<sub>4</sub>C<sub>5</sub>C<sub>7</sub>O<sub>8</sub>) torsion angles for compound (F) shown in Fig. 3.

We have performed *ab initio* solution phase geometry optimizations at the HF/6-31G\*\* level on the structures cor-

TABLE VI. Initial and final solution phase energies  $E_s^i$ ,  $E_s^f$  from Eq. (5.1), solvation energies  $\Delta G^f$  and changes in solvation energies  $\Delta\Delta G^f$  for compounds (A)–(E) resulting from the solution phase geometry optimization calculations.

Molecule	$E_s^i$ [Hr]	$E_s^f$ [Hr]	$\Delta G$ [kcal/mol]	$\Delta\Delta G$ [Hr]/[Kcal/mol]
Acetamide	-208.007 64	-208.008 74	-9.79	-0.001 10/-0.69
<i>N</i> -methylacetamide	-247.031 27	-247.032 30	-11.70	-0.001 03/-0.65
Acrolein	-190.779 59	-190.779 78	-6.34	-0.000 19/-0.12
Mesityl oxide	-307.908 23	-307.908 40	-5.15	-0.000 17/-0.11
Merocyanine	-667.030 86	-667.041 10	-20.04	-0.010 24/-6.43

responding to the various local minima present on the gas phase Ramachandran map. The starting geometries are labeled  $C5$ ,  $C7_{ax}$ ,  $C7_{eq}$ ,  $\alpha_L$ ,  $\alpha_P$ ,  $\beta$ , and  $\beta_2$ . The  $C5$  and  $C7$  structures correspond to extended and cyclic structures of the dipeptide respectively. Those labeled  $\alpha$  and  $\beta$  correspond to helical and strand conformations.

The various structures were initially optimized in the gas phase using starting configurations obtained from the work of Gould and co-workers,<sup>1</sup> except for  $\alpha_P$ , which is not discussed in their study. They also report the presence of an  $\alpha_R$  helical structure minimum in the gas phase diagram corresponding to ( $\Phi = -60.7$   $\Psi = -40.7$ ). We have found no corresponding gas-phase minimum at this location in our calculations both using the PS-GVB<sup>23</sup> and GAUSSIAN92<sup>26</sup> programs. Initial geometries corresponding to the  $\alpha_R$  conformation were found to minimize to the  $\beta_2$  conformation.

The gas-phase minima were subsequently used as starting points for the solution phase geometry optimizations. The locations of the gas and solution phase minima for the different structures are indicated in Fig. 6 and superimposed on the contour lines obtained using the MM3 molecular mechanics potential.<sup>27</sup> As expected *ab initio* minima are either shifted from the molecular mechanics minima or do not have

any corresponding extremum in the classical potential. Table VII list the  $\Psi$  and  $\Phi$  angle structural parameters for the different configurations reproduced from the article by Gould *et al.*,<sup>1</sup> and from our calculations. The gas phase minima for structures  $C_5$  and  $\alpha_R$  disappear in the SCRFF-R solution-phase optimization process ( $\alpha_R$  collapses to  $\beta_2$ ,  $C_5$  collapses to  $\beta$ ). The solvated optimizations performed using our SCRFF-PBF algorithm, by contrast reveal a stable solution phase counterpart to the gas phase  $C_5$ . As pointed out above, the  $\alpha_R$  and  $\alpha_P$  correspond to stable structures appearing only in either method so that little can be said about their relationship. In addition, to these qualitative differences, it is apparent that even when there exists a correspondence between the structures, the two methods produce quantitatively different geometries in the solvated optimization process. The large difference in the  $\Psi$  angle changes observed for the  $\beta$  conformation between the two methods is remarkable.

Tables VIII and IX list the total solution phase and solvation free energies of the various conformations relative to the  $C7_{eq}$  conformation. The corresponding changes in solution phase energies and dipole moments are shown in Table X. A gradient is not yet available for the  $\Delta G^{sa}$  hydrogen-bonding correction term described earlier in this article. The

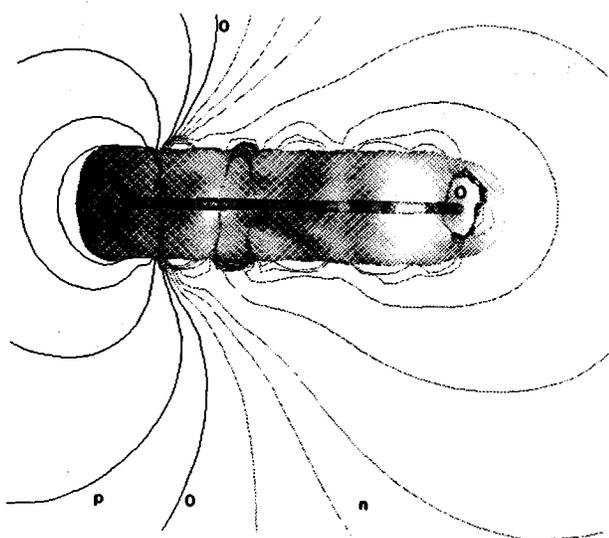


FIG. 4. Cross section of electrostatic potential contour surfaces and dielectric interface for compound (E) in solution. Zero potential surface, positive/negative potential regions, and O,N sites are indicated. Structure is *gas-phase* minimum.

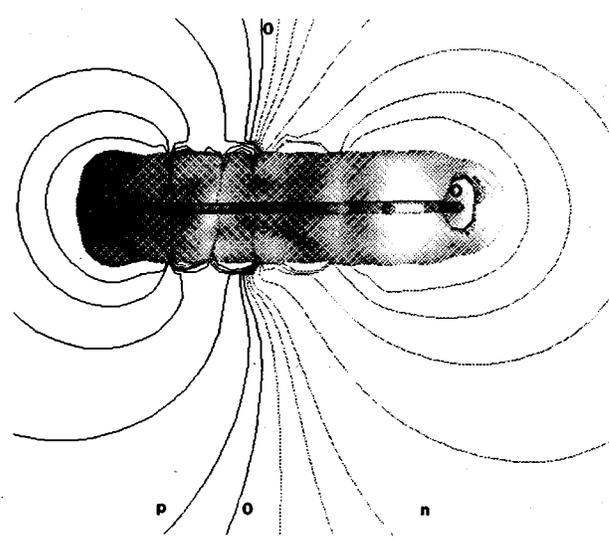


FIG. 5. Cross section of electrostatic potential contour surfaces and dielectric interface for compound (E) in solution. Zero potential surface, positive/negative potential regions, and O,N sites are indicated. Structure is *solution-phase* minimum.

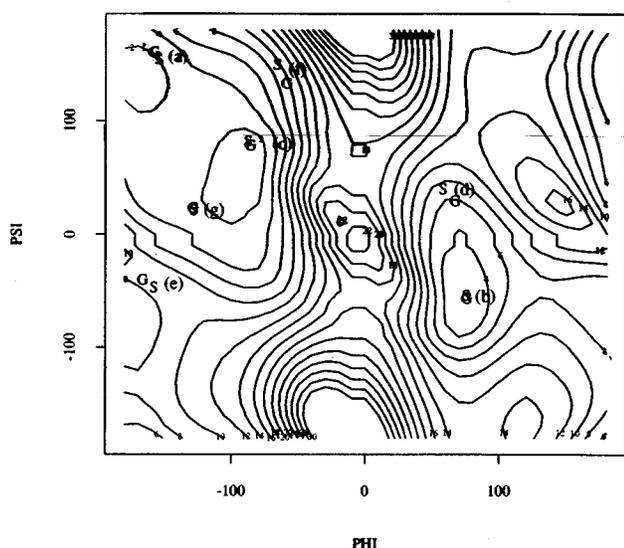


FIG. 6. Contour lines for the MM3 molecular mechanics potential as a function of the  $(\Psi, \Phi)$  torsional angles and location of minima found from *ab initio* calculations at the HF6-31G\*\* level. (G),(S) indicate the location of gas and solution phase minima, respectively, for the seven local minima discussed in text. The regions are (a)  $C5$ , (b)  $C7_{ax}$ , (c)  $C7_{eq}$ , (d)  $\alpha_L$ , (e)  $\alpha_P$ , (f)  $\beta$ , and (g)  $\beta_2$ . Values of contour lines are indicated in [kT] at 298 K.

energies listed in Tables VIII and IX however, include the hydrogen bonding correction term even though its contribution to the total gradient was not included in the minimization process. We have found the ordering of the relative energies to be generally unaffected by this term, since the magnitude of the correction varies little from one conformation to another. There is also little change in the correction between the gas and solution phase optimized structures for each conformation with the exception of the  $\alpha_P$  conformation, where a  $-0.40$  kcal/mol change is observed. The development of a gradient for this energy term is currently under way and will be used to assess the effect of these interactions on the location of the minima.

By comparing the ordering of the structures obtained with our method with the one predicted by the SCRFR method reported by Gould *et al.*<sup>1</sup> we find, in contrast to their results, that the  $\beta$  structure is most favored in solution, instead of  $\beta_2$  (see Table VIII). The PCM energies for the

TABLE VIII. Total solution phase energies of the AD conformations. (G) Indicates a value for the solvated gas phase optimized geometry, (S) a value for the solution phase optimized geometry. All energies are relative to the  $C7_{eq}$  conformation. Results are shown for (A) SCRFR, (B) PCM, and (C) SCRFR-PBF methods. The PCM calculations are for the SCRFR solvated optimized geometries. For SCRFR-PBF energies the hydrogen bonding correction factor discussed in text is included in the tabulated energies but not in the optimization process. All energies are in [kcal/mol].

Conformation	$\Delta E_s - G$	$\Delta E_s - S$	$\Delta E_s - G$	$\Delta E_s - S$	$\Delta E_s - G$	$\Delta E_s - S$
	A (rel)	A (rel)	B (rel)	B (rel)	C (rel)	C (rel)
$C7_{eq}$	0.00	0.00	0.00	0.00	0.00	0.00
$C7_{ax}$	1.69	0.07	2.97	-0.66	2.40	2.59
$C5$	-0.11	...	-0.73	...	-0.81	-0.97
$\alpha_R$	0.10	...	1.59	...	...	...
$\alpha_P$	...	...	...	...	5.06	3.18
$\alpha_L$	3.52	1.89	2.55	0.69	1.67	1.07
$\beta$	5.68	-4.27	4.82	-3.59	0.75	-1.25
$\beta_2$	-2.98	-5.44	0.65	-1.47	0.27	-0.41

SCRFR geometries however are in slightly better agreement with our calculations and also predict the  $\beta$  structure to be the most stable. For all the structures except the  $\beta$  case, this can probably be attributed to the fact that the PCM model uses a more realistic cavity in the dielectric continuum calculations than the SCRFR model so that for similar geometries, the energy calculations between the two methods should be consistent. For the  $\beta$  confirmation the differences between the SCRFR and SCRFR-PBF structures suggest that the reasons for the improved agreement are not as clear. A more precise comparison is difficult to make at this point since the reference point used in Gould's and our calculations ( $C7_{eq}$ ) differ somewhat in the solution phase  $(\Phi, \Psi)$  coordinates due to the use of very different cavity shapes in the solvated optimization procedure.

Given that the gas phase optimized structures for all conformations appear to be structurally similar in Gould's and our calculations, the large energetic difference observed for the solvated  $\beta$  conformation remains unexplained. Since the solvation free energies obtained with the PCM and SCRFR-PBF methods listed in Table IX are in better agreement than the total free energies, the discrepancy probably lies in the gas phase energy of the conformation.

It is unclear at present whether it is possible at all, using NMR techniques, to identify with certainty which of the con-

TABLE VII. Structural parameters for various gas (G) and solution phase (S) optimized conformations of the alanine dipeptide.

Conformation	$\Phi - G^a$	$\Psi - G^a$	$\Phi - S^a$	$\Psi - S^a$	$\Phi - G^b$	$\Psi - G^b$	$\Phi - S^b$	$\Psi - S^b$
$C7_{eq}$	-85.8	79.0	-73.4	75.1	-85.80	78.49	-87.48	82.07
$C7_{ax}$	76.0	-55.4	74.9	-73.4	75.84	-56.50	75.33	-53.80
$C5$	-152.7	159.8	...	...	-157.86	160.30	-154.48	154.60
$\alpha_R$	-60.7	-40.7	...	...	...	...	...	...
$\alpha_P$	...	...	...	...	-166.40	-40.11	-157.84	-46.71
$\alpha_L$	67.0	30.2	68.4	39.3	66.87	29.51	58.04	40.18
$\beta$	-57.6	134.4	-118.2	133.1	-58.90	133.30	-65.83	149.51
$\beta_2$	-130.9	22.3	-112.1	22.5	-128.59	23.24	-127.70	21.48

<sup>a</sup>Results from Gould and co-workers (Ref. 1).

<sup>b</sup>Results of the SCRFR-PBF calculations. A (...) indicates that no corresponding structure was found. All angles shown in degrees.

TABLE IX. Solvation free energies of the AD conformations. (*G*) Indicates a value for the solvated gas phase optimized geometry, (*S*) a value for the solution phase optimized geometry. All energies are relative to the  $C7_{eq}$  conformation. Results are shown for (A) SCRF-R, (B) PCM, and (C) SCRF-PBF methods. The PCM calculations are for the SCRF-R solvated optimized geometries. For SCRF-PBF energies the hydrogen bonding correction factor discussed in text is included in the tabulated energies but not in the optimization process. All energies are in [kcal/mol].

Conformation	$\Delta G-G$	$\Delta G-S$	$\Delta G-G$	$\Delta G-S$	$\Delta G-G$	$\Delta G-S$
	A (rel)	A (rel)	B (rel)	B (rel)	C (rel)	C (rel)
$C7_{eq}$	0.00	0.00	0.00	0.00	0.00	0.00
$C7_{ax}$	-1.33	0.07	-0.05	-0.66	-0.51	-0.32
$C5$	0.24	...	-0.38	...	-1.15	-1.30
$\alpha_R$	-4.40	...	-2.91	...	...	...
$\alpha_P$	...	...	...	...	-0.68	-2.56
$\alpha_L$	-1.63	-1.17	-2.60	-2.37	-3.09	-3.70
$\beta$	0.32	-3.49	-0.54	-2.99	-1.08	-3.08
$\beta_2$	-4.86	-5.79	-1.23	-1.82	-2.23	-2.91

formations is favored in solution. Indeed, it is necessary to construct a model for the experimental data in order to extract a possible ordering for the structures. To this extent, the identification of the  $\beta$  structure as the most stable is in agreement with the model proposed by Madison and Kopple<sup>28</sup> based on their experimental measurements. The raw data itself, however, does not appear unambiguous so that the correlation between the result of the *ab initio* calculations and the model could be fortuitous. It should be also noted, that both Gould's and our calculations fail to predict the existence of the  $\alpha_R$  conformation in solution which Madison and Kopple claim to have identified.

Not surprisingly, we observe that for most conformations, the changes in the total free energy in the optimization process, correlate with the change in the solvated dipole moment of the molecule. The dipole moments listed were obtained from the ESP fitted partial charge representation of the molecule, and formally differ from the dipole moments obtained from the complete quantum mechanical charge distribution. Calculations done in our laboratory however indicate that in practice the two dipole moments differ negligibly ( $\sim$  a few percent) when the RMS error in the ESP fitting process is small. The  $C7_{eq}$  structure appears to be an exception though as the dipole moment decreases as a result of the optimization process.

TABLE X. Changes in total solution phase energies of the AD conformations and dipole moments from ESP fitted charges resulting from geometry optimization using the SCRF-PBF method. All energies are in [kcal/mol] and dipoles in [D] units.

Conformation	$\Delta \Delta E_s$	$\mu_G$	$\mu_S$
$C7_{eq}$	-0.71	3.4	3.1
$C7_{ax}$	-0.52	6.7	7.0
$C5$	-0.86	3.9	4.2
$\alpha_P$	-2.59	3.7	4.0
$\alpha_L$	-1.31	6.2	6.2
$\beta$	-2.71	3.2	5.6
$\beta_2$	-1.39	7.0	7.6

TABLE XI. Number of geometry optimization steps and total CPU time required to obtain solution phase optimized geometries for the different alanine dipeptide conformations. Calculations were performed on IBM390 workstations.

Conformation	$N_{iter}^s$	$T$ [hr:min]
$C7_{eq}$	26	9:04
$C7_{ax}$	9	3:13
$C5$	21	7:19
$\alpha_P$	12	4:16
$\alpha_L$	27	9:31
$\beta$	22	7:47
$\beta_2$	24	8:22

While for the  $\beta$  and  $\beta_2$  structures the large changes in energies and dipole moments appear to correlate with the structural changes listed in Table VII, this does not seem to hold for the  $\alpha_L$  structure in the light of large torsion angle changes resulting in only moderate energy changes and no change in the molecular dipole moment. This unusual situation does not appear to be connected to the relative measurement since the value of  $\Delta \Delta E_s$  ( $\Delta \Delta G$ ) indicates that there is a net stabilization of 1.31 kcal/mol due to the solvent.

Finally, the total CPU time and number of solution phase geometry optimization iterations required to obtain the converged geometries, starting from the gas phase optimized structures, is shown in Table XI. Calculations were performed on the IBM390 workstations available at the Columbia SP2 parallel computing facility. The average CPU time per optimization iteration step required was approximately 1250 CPU seconds, making the solution phase optimization procedure about 1.4 times more expensive than its gas phase counterpart.

## VII. CONCLUSION

We have presented a formalism and method for computing solvation free energy gradients and carrying out solution phase geometry optimizations within the SCRF method. The algorithm is based on our existing three-dimensional finite element formulation of the continuum electrostatics problem which will be detailed in an upcoming publication. Several tests have indicated the accuracy of the gradient components obtained from the method and have validated the proposed strategy for extending the SCRF method developed in our laboratory<sup>5</sup> to carry out solution phase geometry optimizations. Applications of the method to a number of test molecules indicate that, for rigid compounds the primary effect of the solvent on the structure is to amplify the charge separation process, which can be observed even if identical structures are placed in different media, by altering the bonding nature of the molecule. For more flexible systems, such as the alanine dipeptide molecule discussed above, the structural changes observed are correlated with the change in the total dipole moment of the molecule. A more extensive study of the performance of the method, as well as of the effects of higher level electron correlations using polyalanine systems, is currently under way. Preliminary results using LMP2 en-

ergy corrections on this system indicate that the total free energy of the conformations at the solution phase minima is systematically decreased by the addition of the perturbation term. Work is also under way to integrate a different gradient algorithm based on Gilson's formalism with molecular mechanics methods to investigate solvation effects within the dielectric continuum model on larger structures.

## ACKNOWLEDGMENTS

We would like to thank Dr. Robert Murphy and Dr. Murco Ringnalda for their help with the integration of the PBF and PS-GVB programs. Also, we thank Dr. Marshall Newton for many helpful discussions and suggestions regarding the SCRF formalism. This work was supported in part by grant R01 GM40526 from the NIH. C.M.C. acknowledges support from the FCAR fund.

- <sup>1</sup>Ian R. Gould, W. D. Cornell, and I. H. Hillier, *J. Am. Chem. Soc.* **116**, 9250 (1994).
- <sup>2</sup>J. A. Grant, R. L. Williams, and H. A. Scheraga, *Biopolymers* **30**, 929 (1990).
- <sup>3</sup>J. Tomasi, R. Alagona, R. Bonaccorsi, and C. Ghio, in *Modeling of Structures and Properties of Molecules*, edited by C. Maksic (Eliis Horwood, Cichester, U.K., 1987), p. 330.
- <sup>4</sup>J. Bajorath, J. Kraut, Z. Li, D. H. Kitson, and A. T. Hagler, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6423 (1991).
- <sup>5</sup>D. J. Tannor, B. Marten, R. Murphy, R. A. Friesner, D. Sitkoff, A. Nicholls, M. Ringnalda, W. A. Goddard III, and B. Honig, *J. Am. Chem. Soc.* **116**, 11875 (1994).
- <sup>6</sup>B. Marten, K. Kim, C. M. Cortis, R. A. Friesner, R. B. Murphy, M. N. Ringnalda, D. Sitkoff, and B. Honig, *J. Phys. Chem.* **100**, 11775 (1996).
- <sup>7</sup>J. L. Rivail and B. Terryn, *J. Chem. Phys.* **79**, 2 (1982).
- <sup>8</sup>D. Rinaldi, *Comput. Chem.* **6**, 155 (1982).
- <sup>9</sup>D. Rinaldi, J. L. Rivail, and N. Rguini, *J. Comput. Chem.* **13**, 675 (1992).
- <sup>10</sup>R. Cammi and J. Tomasi, *J. Chem. Phys.* **100**, 7495 (1994).
- <sup>11</sup>R. Cammi and J. Tomasi, *J. Chem. Phys.* **101**, 3888 (1994).
- <sup>12</sup>C. Cortis and R. Friesner (in preparation).
- <sup>13</sup>K. Sharp and B. Honig, *Annu. Rev. Biophys. Biophys. Chem.* **19**, 301 (1990).
- <sup>14</sup>J. Tomasi and M. Persico, *Chem. Rev.* **94**, 2017 (1994).
- <sup>15</sup>A. Ben-Naim and Y. Marcus, *J. Chem. Phys.* **81**, 2016 (1984).
- <sup>16</sup>I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig, *Proteins* **1**, 47 (1986).
- <sup>17</sup>Y. N. Vorobjev, J. A. Grant, and J. A. Scheraga, *J. Am. Chem. Soc.* **114**, 3189 (1992).
- <sup>18</sup>F. M. Richards, *Annu. Rev. Biophys. Bioeng.* **6**, 151 (1977).
- <sup>19</sup>T. J. Baker, *Appl. Numer. Math.* **5**, 275 (1989).
- <sup>20</sup>M. Gilson, M. E. Davis, B. A. Luty, and J. A. McCammon, *J. Phys. Chem.* **97**, 3591 (1993).
- <sup>21</sup>J. F. Gaw, Y. Yamaguchi, H. F. Scheafer III, and N. C. Handy, *J. Chem. Phys.* **85**, 5132 (1986).
- <sup>22</sup>Y. Won, J.-G. Lee, M. N. Ringnalda, and R. A. Friesner, *J. Chem. Phys.* **94**, 8152 (1991).
- <sup>23</sup>M. N. Ringnalda, J.-M. Langlois, B. H. Greeley, R. B. Murphy, T. V. Russo, C. Cortis, R. P. Muller, B. Marten, R. E. Donnelly Jr., D. T. Mainz, J. R. Wright, W. T. Pollard, Y. Cao, Y. Won, G. H. Miller, W. A. Goddard III, and R. A. Friesner, PS-GVB v2.1. (Schrödinger, Inc., 1995).
- <sup>24</sup>T. Head-Gordon, M. Head-Gordon, M. J. Frisch, C. L. Brooks III, and J. A. Pople, *J. Am. Chem. Soc.* **113**, 5989 (1991).
- <sup>25</sup>S. Miertus, E. Scrocco, and J. Tomasi, *J. Chem. Phys.* **55**, 117 (1981).
- <sup>26</sup>GAUSSIAN 92, Revision E.1., M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople (Gaussian, Inc., 1992).
- <sup>27</sup>N. Allinger and L. Yan, *J. Am. Chem. Soc.* **115**, 11918 (1993).
- <sup>28</sup>V. Madison and K. D. Kopple, *J. Am. Chem. Soc.* **102**, 4855 (180).