Geometry optimization of molecules in solution: Joint use of the mean field approximation and the free-energy gradient method

I. Fdez. Galván, M. L. Sánchez, M. E. Martín, F. J. Olivares del Valle, and M. A. Aguilar
Departamento Quimica-Fisica, Universidad de Extremadura, Avda de Elvas s/n, 06071 Badajoz, Spain

(Received 17 June 2002; accepted 7 October 2002)

The average solvent electrostatic potential/molecular dynamics (ASEP/MD) and the free-energy gradient methods are applied together with the multidimensional geometry optimization of molecules *in solution*. The systems studied were formamide in aqueous solution and water and methanol in liquid phase. The solute molecules were described through *ab initio* quantum mechanics methods (density dunctional theory or Møller–Plesset second order perturbation theory) while the solvent structure was obtained from Molecular Dynamics calculations. The method is very efficient; the increase in computation time is minimal with respect to previous ASEP/MD versions that worked at a fixed geometry. Despite the use of the mean field approximation in the calculation of the solvent reaction potential the agreement with previous theoretical calculations was satisfactory. Large changes were observed in the solute charge distribution induced by the solvent, and the solute polarization was accompanied by an increase in the solvent structure around the solute. © *2003 American Institute of Physics.* [DOI: 10.1063/1.1525798]

I. INTRODUCTION

The optimization of structures (minima or saddle points) *in solution* is a complicated problem because of the difficulty in properly accounting for the influence of the bulk solvent polarization and the thermal effects. In the case of the bulk solvent polarization, especially in polar liquids, it is compulsory to consider a great number of solvent molecules so as to adequately describe the electric field generated by the solvent in the volume occupied by the solute, while to include thermal effects a great number of configurations have to be considered, with the computational cost that these two factors imply. One way to simplify the problem is by having recourse to the Mean Field Approximation (MFA).[1,2] The introduction of the MFA has been a constant aspect of the development of solvent effect theories. The different quantum versions of continuum models,[3] for instance, make use of this approximation. More recently, several methods[4,5] have been proposed that combine MFA with a detailed description of the solvent structure obtained from an extended version of the reference interaction site model[6] or from Molecular Dynamics simulations.[7]

In previous papers[5] our group has developed a nonstandard Quantum Mechanics/Molecular Mechanics method (QM/MM) that makes use of the MFA. The method, denoted ASEP/MD, is based on the introduction of the Averaged Solvent Electrostatic Potential obtained from Molecular Dynamics simulation into the solute molecular Hamiltonian. As in other mean field theories the mean value of the energies of the different configurations is replaced by the energy of an average configuration that in our case is obtained from MD calculations. Its main advantage is that it permits reducing the number of quantum calculations from several thousand to 4–8 without introducing significant errors either in the energy or the molecular properties. The method has been successfully applied to the study of liquids[5(e),5(f)] and to the determination of solvent shifts in the UV/VIS spectra.[5(b)–5(d)].

Several strategies can be used in QM/MM calculations to obtain optimal structures *in solution*. In traditional QM/MM methods,[8] for instance, the internal degrees of freedom are considered to be additional dynamical variables subject to thermal fluctuations. At each step, the total force (solute+solute contributions) over the solute atoms is calculated and a new geometry is obtained by integration of the equations of motion. A mean geometry can then be obtained by averaging over the different solute structures.

In a second set of methods, known as free-energy gradient methods[9–11] the forces felt by the solute atoms are obtained from QM/MM simulations where the solute molecule has a fixed structure. The free-energy surface (FES) is defined as the time average of the forces acting on each atom of a solute molecule over the equilibrium distribution for all solvent molecules.[10] From the mean gradient, a new geometry can be generated and the process is repeated until the gradient converges to a desired precision. While the main advantage of this method is that it permits obtaining both stable and transition states, it is at the expense of notably increasing the simulation time and hence the computational cost with respect to traditional QM/MM.

Because of its use of average quantities, the free-energy gradient method seems especially adequate to use together with the MFA. Their joint application permits a considerable saving of computation time. In what follows, we present a method of obtaining geometries and properties *in solution* that combines the FEG method with ASEP/MD. The method permits obtaining both stable and transition states. Because of the small number of quantum calculations that it involves, the solute description can be performed at exactly the same level as is used for *in vacuo* calculations. In this work we

 © 2003 American Institute of Physics

restrict ourselves to the application of the ASEP/MD model to the problem of obtaining optimized structures *in solution*. We postpone the more complicated case of searching for transition states to a future paper.

The rest of the paper is organized as follows: In Sec. II the computational method is explained. Section III provides a technical details; specifically, we discuss the nature and magnitude of the gradient fluctuations and how they affect the optimization procedure. The results are discussed in Sec. IV. The final section presents the main conclusions.

II. METHOD

The main characteristic of the ASEP/MD has been discussed elsewhere.[5] Here, we center on the special features of its application to the determination of critical points on potential energy surfaces.

As in traditional QM/MM methods,[12] in ASEP/MD, the energy and state function of the solvated solute molecules are obtained by solving the effective Schrödinger equation,

$$(\hat{H}_{\mathrm{QM}}+\hat{H}_{\mathrm{QM/MM}}|\Psi=E|\Psi. \tag{1}$$

The interaction term, $\hat{H}_{\mathrm{QM/MM}}$ takes the following form:

$$\hat{H}_{\mathrm{QM/MM}}=\hat{H}^{\mathrm{elect}}_{\mathrm{QM/MM}}+\hat{H}^{\mathrm{vdw}}_{\mathrm{QM/MM}}, \tag{2}$$

$$\hat{H}^{\mathrm{elect}}_{\mathrm{QM/MM}}=\int dr\cdot\hat{\rho}\cdot\langle\hat{V}_s(r;\rho)\rangle, \tag{3}$$

where $\hat{\rho}$ is the solute charge density and the brackets denote a statistical average. The term $\langle\hat{V}_s(r;\rho)\rangle$ is the average electrostatic potential generated by the solvent at the position r,and is obtained from MD calculations where the solute molecule is represented by the charge distribution ρ and a geometry fixed during the simulation. The term $\hat{H}^{\mathrm{vdw}}_{\mathrm{QM/MM}}$ is the Hamiltonian for the van der Waals interaction, in general represented by a Lennard-Jones potential. Given that the solvent structure, and hence the ASEP, is a function of the solute charge density, Eqs. (1) and (3) have to be solved iteratively. In general, only a few cycles of quantum calculation/molecular dynamics simulations are needed for convergence.

Next, we shall describe the application of the MFA to the determination of the gradient and Hessian. By way of comparison, and to determine the nature of the approximation introduced, we first present their expressions when the MFA is not used. In this case the force on the free-energy surface (FES) is[9–11]

$$F(r)=-\frac{\partial G(r)}{\partial r}=-\left\langle\frac{\partial V(r)}{\partial r}\right\rangle, \tag{4}$$

where $G(r)$ is the free-energy, V is the sum of the contributions associated with the interaction with the other atoms of the solute molecule, V_i, and with the solute–solvent interaction energy, V_s and the brackets denote a statistical average.

The Hessian is

$$H\left\langle\frac{\partial^2 V}{\partial r\partial r}\right\rangle-\beta\left\langle\frac{\partial V}{\partial r}\frac{\partial V^T}{\partial r}\right\rangle+\beta\left\langle\frac{\partial V}{\partial r}\right\rangle\left\langle\frac{\partial V}{\partial r}\right\rangle^T, \tag{5}$$

$$H=\left\langle\frac{\partial^2 V}{\partial r\partial r}\right\rangle-\beta[\langle F^2\rangle-\beta\rangle F\rangle^2], \tag{6}$$

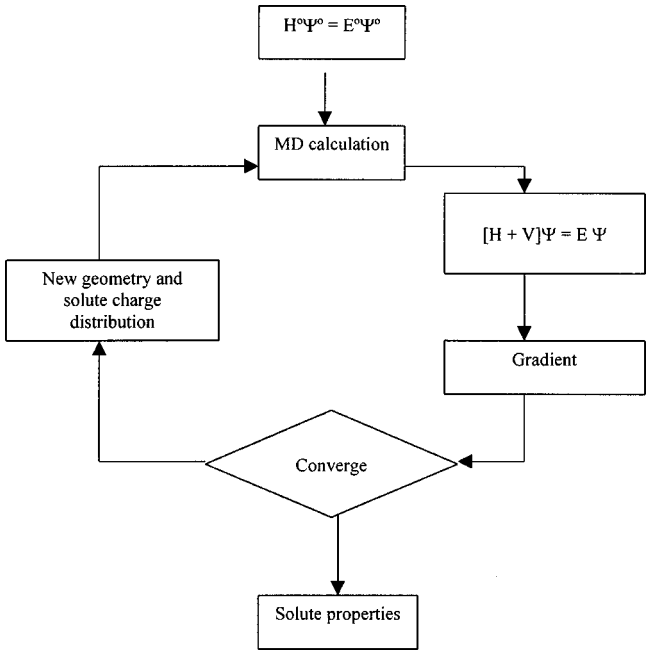

FIG. 1. Geometry optimization scheme.

where the superscript T denotes the transposition and $\beta=1/RT$. The last term in Eq. (6) is related to the thermal fluctuation of the force.

Now we shall analyze how we can obtain the force and the Hessian in ASEP/MD.

Because we assume a fixed geometry and a fixed charge distribution of the solute during the simulation, the average value of the force can be replaced by the force of the mean configuration,

$$F(r)=-\frac{\partial\langle V\rangle}{\partial r}, \tag{7}$$

$$H=\frac{\partial^2\langle V\rangle}{\partial r\partial r}-\beta\left\langle\frac{\partial V}{\partial r}\frac{\partial V^T}{\partial r}\right\rangle-\beta\frac{\partial\langle V\rangle}{\partial r}\frac{\partial\langle V\rangle^T}{\partial r}. \tag{8}$$

In this point we introduce an additional approximation: we neglect the force fluctuations. Given that the Hessian is used only to accelerate the optimization procedure, this approximation has no effect on the optimized geometries. In any case, preliminary estimations show that the errors introduced in the trace of the Hessian in the formamide–water systems when we neglect the fluctuation term is lower than 5%. With this approximation the Hessian reads

$$H=\frac{\partial^2\langle V\rangle}{\partial r\partial r}. \tag{9}$$

Equations (7) and (9) are the equations traditionally used by mean field theories[3,13] and are the ones that we will follow in the present paper.

Figure 1 shows the scheme of the optimization procedure with ASEP/MD. The procedure is as follows:

(1) We begin by obtaining through an *ab initio* calculation the *in vacuo* solute charge distribution that is used as input in the molecular dynamics calculation in the next step;

(2) From the MD data, we obtain the ASEP, $\langle \hat{V}_s(r) \rangle$, and the solvent contribution to the gradient and Hessian;
(3) The ASEP is introduced into the molecular Hamiltonian of the solute. The electronic wave function of the solute, now *in solution*, can be obtained by solving the associated effective Schrödinger equation;
(4) The solute contributions to the gradient and Hessian are calculated and added to the solvent contribution. A new geometry is obtained;
(5) With the new geometry the electronic wave function of the solute *in solution* is obtained and a new solute charge distribution is calculated. The new geometry and solute charge distribution are then used as an input in a new molecular dynamics calculation [step (2)]. This process is repeated until convergence in the free-energy is reached.

At each step of the self-consistent process, the solute charges used in the MD calculation were obtained by fitting the molecular electrostatic potential of the solute molecule in the presence of the solvent perturbation in the standard way. The CHELP program was used.[14]

The procedure for obtaining the new geometry from the old one, step (4), deserves more attention. We have checked three possibilities:

(1) Using only the force to determine the position of the next point on the free-energy surface

$$q_{k+1}=q_k+F_k. \tag{10}$$

(2) Finding the following point using the force and the Hessian,

$$q_{k+1}=q_k+H_k^{-1}F_k. \tag{11}$$

(3) Performing the complete optimization of the solute geometry at each step of the cycle. This last option is the most expensive computationally, and does not seem worthwhile, especially if one takes into account that the optimized geometry will be distorted in the next cycle due to the fluctuation in the gradient.

In the three cases the forces and Hessians are calculated analytically. Compared with traditional QM/MM models the errors introduced by the above procedure are the following:

(1) The solute charge distribution obtained by solving the Schrödinger equation in the presence of a mean perturbation, (ρ), is different from the averaged value of the solute charge distribution obtained for each solvent configuration, i.e., $\overline{\rho} \neq \langle \rho \rangle$. This difference appears because in traditional QM/MM methods the solute charge distribution is fitted at each step to the new solvent configuration, which is not permitted in MF theories. Given that $\langle V_i \rangle$ and $\langle V_s \rangle$ are both functions of ρ, this approximation affects the energy, the gradient, and the Hessian that are not calculated with the correct solute charge distribution. The magnitude of the errors introduced by the MFA in the calculation of the energy and the solute charge distribution has been addressed in previous papers,[2,5] and was evaluated to be $<5\%$ for the energy and 1% for the dipole moment. The magnitude of the errors introduced by the MFA in the magnitude of the gradient is discussed below.

(2) In the calculation of the Hessian, the fluctuations of the force, the last term in Eq. (6), are neglected. Given that the Hessian is used only to accelerate the optimization procedure, this approximation, in principle, has not effect on the optimization of the geometry, and will not be considered in this paper.

III. TECHNICAL DETAILS

In this paper, the model described above is applied to the study of several molecules in the condensed phase. We studied both liquids and solutions, and employ different levels of calculation in order to check the performance of the method in different situations. As an example of a solution, we took formamide in an aqueous solution. The basis set[15] used was NCO(711/411/1)H(41/1) and the solute wave function was obtained with DFT techniques. During the DFT calculation the density-gradient-corrected correlation functional proposed by Perdew[16] and Becke's[17] exchange functional were used.

As examples of liquids, we took water and methanol. In both cases, the basis set used was the aug-cc-pVDZ from Dunning *et al.*[18] The level of calculation was the second-order Møller–Plesset (MP2) perturbation theory for the water and DFT for the methanol. The functional was the same as used for formamide.

In the three systems the initial solute geometry was that optimized in the gas phase at the level and for the basis sets indicated above. However, in each case a different procedure was used to obtain the next solute geometry. In formamide, the geometry was completely optimized at each cycle of the ASEP/MD procedure. In methanol, the gradient and Hessian were used to determine the next geometry. Lastly, in water, only the gradient was used. All the quantum calculations were done using the GAUSSIAN98 package.[19]

The MD calculations were performed using the program MOLDY.[20] In each case, 128 (methanol) or 215 (water and formamide) molecules were simulated at a fixed intramolecular geometry by combining Lennard-Jones interatomic interactions with electrostatic interactions. In the formamide–water and water–water systems, the 214 solvent molecules were simulated by the TIP3P (Ref. 21) model at a fixed intramolecular geometry. The formamide–water potential parameters were taken from Jorgensen and Swenson.[22] The geometry and parameters of the classical methanol molecules were taken from Jorgensen *et al.*[23] Periodic boundary conditions were applied, and spherical cutoffs were used to truncate the molecular interactions at 9.0 Å. A time step of 0.5 fs was used. The electrostatic interaction was calculated with the Ewald method.[22] The temperature was fixed at 298 K by using the Nosé–Hoover[24] thermostat. Except in the cases expressly indicated, each MD calculation simulation was run for 150 000 time steps (50 000 equilibration, 100 000 production).

TABLE I. Mean value and standard deviation (in 10^{-3} hartree/bohr) of the total Cartesian free energy gradient of formamide in aqueous solution, as well as their respective rms. The ASEP was calculated every 50 or 100 ps and the gradient was obtained.

		50 ps		100 ps	
		Mean	Std. dev.	Mean	Std. dev.
	x	−0.765	1.154	−0.706	0.173
N1	*y*	−0.892	0.372	−0.798	0.262
	z	−0.237	0.367	−0.212	0.045
	x	−0.054	0.625	−0.084	0.175
H2	*y*	0.616	0.348	0.632	0.132
	z	0.182	0.342	0.151	0.098
	x	−0.316	0.169	−0.288	.097
H3	*y*	−0.424	0.545	−0.498	0.343
	z	0.110	0.141	0.101	0.074
	x	−3.009	2.004	−3.040	0.261
C4	*y*	−0.682	0.557	−0.662	0.273
	z	0.003	0.046	−0.011	0.028
	x	−1.820	1.510	−1.695	0.496
O5	*y*	−1.389	0.727	−1.400	0.179
	z	0.164	0.105	0.245	0.136
	x	−0.054	0.063	−0.051	0.048
H6	*y*	0.102	0.135	0.107	0.036
	z	−0.002	0.063	0.000	0.007
rms		0.970	0.736	0.959	0.202

IV. RESULTS AND DISCUSSION

Prior to any geometry optimization, there are several numerical questions to consider. First, we have to determine the criterion of convergence of the gradient for optimizations *in solution* using the FEG method. The problem arises because, *in solution*, to determine the solvent contribution to the gradient we need to perform a thermodynamic average. Given the finite time of the simulations, calculations performed under the same conditions but starting from different points of the configuration space will yield different values of the gradient (and any other property, obviously). As a consequence, *in solution*, the precision of a geometry optimization is limited by the gradient fluctuations associated with the finite time of the simulation. Table I gives the mean value of the gradient on the free-energy surface of formamide in aqueous solution, and the magnitude of the largest standard deviation of the gradient (lsdg) as a function of the length of the simulation for a nonoptimized geometry. The root mean square (rms) gradient is also shown. To obtain these values we performed a MD calculation of 300 ps with a time step of 0.5 fs and fixed solute geometry. Next, averages of the gradient were calculated making use of the MFA at intervals of 50 ps and 100 ps. The lsdg can be taken as a measure of the precision at which the optimization can be performed. With 50 ps simulations, the lsdg is 0.0020 hartree/bohr. This value reduces to 0.0005 hartree/bohr when we double the simulation time. Figure 2 shows the molecular Cartesian frame of the formamide molecule.

Next, we shall analyze the magnitude of the errors that the MFA introduces in the gradient evaluation. Table II compares the gradient values for formamide in aqueous solution when the MFA is used and when it is not. From a 75 (25 +50) ps simulation, we chose 1000 configurations. For each

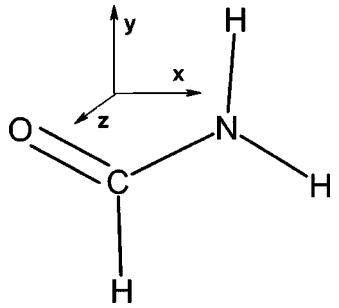

FIG. 2. Molecular Cartesian frame of formamide molecule.

configuration the gradient is obtained, and then the average value. This value is compared with the result obtained when the MFA is used. In this case, from the same set of 1000 configurations we calculate the ASEP, that is introduced into the solute molecular Hamiltonian from which the gradient can be obtained. The magnitude of the largest error is 0.000 23 hartree/bohr, representing a relative error close to 3% with respect to the mean gradient. For 75 ps simulations, this error is an order of magnitude lower than the gradient fluctuations associated with the finite size of the simulations (0.0020 hartree/bohr), and we can expect that the error in the gradient had a negligible effect on the optimized geometry. However, for 150 ps simulations or longer the error associated with the MFA can become dominant, and sets a limit to the precision at which an optimization can be performed when the MFA is used.

Next, we shall consider the results for the optimization of several molecules *in solution*. The first system we studied was formamide in aqueous solution. We chose this system because the *in solution* geometry for this system had previously been considered by different authors using both continuum models and QM/MM methods.[8] Furthermore, it can be seen as the simplest example of a peptide bond. To check

TABLE II. Total Cartesian gradient of the free energy (in 10^{-3} hartree/bohr) of a molecule of formamide in aqueous solution. The Lennard-Jones contribution was not included.

		Mean of 1000 configurations	Average configuration
	x	12.914	12.827
N1	*y*	3.858	3.831
	z	0.084	0.085
	x	−3.365	−3.132
H2	*y*	3.049	2.928
	z	0.053	0.049
	x	−1.116	−1.081
H3	*y*	−5.347	−5.081
	z	−0.006	−0.004
	x	−26.864	−26.816
C4	*y*	5.473	5.487
	z	−0.002	−0.010
	x	18.255	17.862
O5	*y*	−10.052	−9.788
	z	−0.041	−0.026
	x	1.567	1.571
H6	*y*	−2.136	−2.240
	z	−0.004	0.002
rms		8.898	8.807

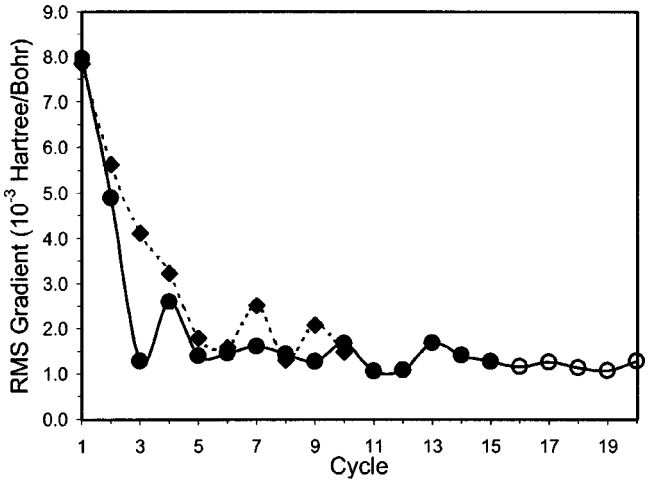

FIG. 3. Root mean square change in free-energy surface of formamide in aqueous solution. 50 ps. (filled circles), 150 ps. (open circles), and 25 ps. (dotted line, diamonds) production simulations.

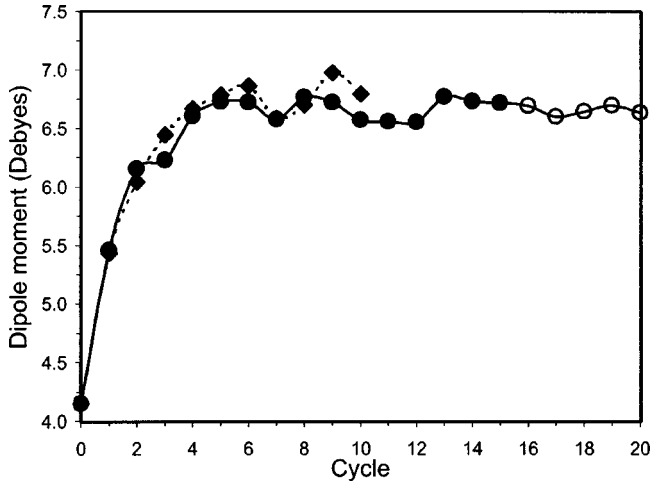

FIG. 5. Dipole moment change of formamide in aqueous solution during the optimization procedure. Symbols: see Fig. 3.

the effect of the gradient fluctuation on the optimization procedure, we performed many more cycles than strictly necessary. Figure 3 displays the root mean square change of the gradient on the free-energy surface during the optimization of formamide in aqueous solution obtained with 75 (25+50) ps simulations. Convergence is reached in 5–6 cycles. From this point, the gradient begins to fluctuate around a rms gradient value of about 0.0015 hartree/bohr. This value is somewhat lower than the values obtained by other authors who also used the FEG method. For instance, Okuyama-Yoshida *et al.*[10] in the optimization of glycine obtain 0.0025 hartree/bohr, and Hirao *et al.*[11] in the transition state of the Menshutkin reaction between ammonia and methyl chloride obtain 0.010 hartree/bohr. The larger values obtained by these authors are probably related to the shorter duration of the simulations, 15+10 ps and 10+10 ps, respectively. Our gradient value is one order of magnitude larger than the thresholds used for *in vacuo* calculations (in Gaussian,[19] for instance, the maximum and rms gradients are 0.000 45 and 0.000 3 hartree/bohr, respectively). Contrary to what was expected, the rms gradient value, Fig. 3, is not reduced when the size of the simulation is increased. Only its fluctuations decrease. This fact is probably an artifact of the ASEP/MD method, related to the different representations of the solute charge distribution during the simulation (classical, through the point charges) and the optimization (quantum, through the wave function).

Figure 4 shows the evolution of the solute polarization free-energy during the optimization procedure. As before, convergence is reached in 5–6 cycles, when ΔG begins to fluctuate. The free energies were calculated by the free-energy perturbation method.[25] As expected, the size of the fluctuations decreases with increasing length of the simulation time, but the average value is almost the same in all the cases (-7.1 ± 1.4 kcal/mol, -7.0 ± 0.7 kcal/mol, and -7.1 ± 0.4 kcal/mol for the 25, 50, and 150 ps simulations, respectively). Only for the longer simulations are the fluctuation less than the fluctuation due to thermal effects, k_BT, which is of the order of 0.6 kcal/mol at 300 K. The same oscillatory behavior is observed for the dipole moment, see Fig. 5. The averaged induced solute dipole moment is 2.5 D, represent-

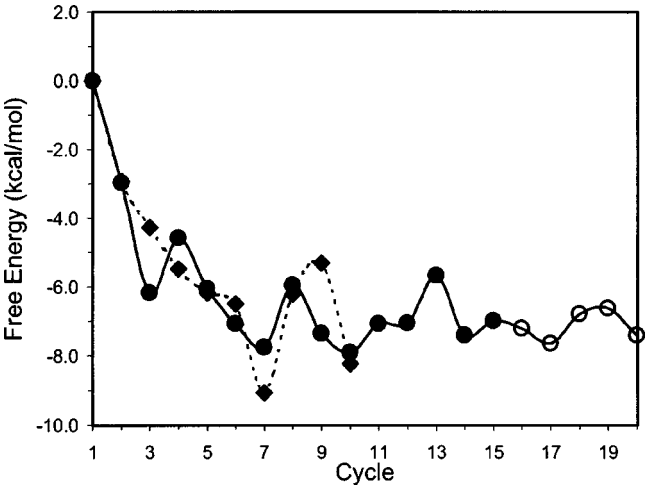

FIG. 4. Polarization free-energy change of formamide in aqueous solution during the optimization procedure. Symbols: see Fig. 3.

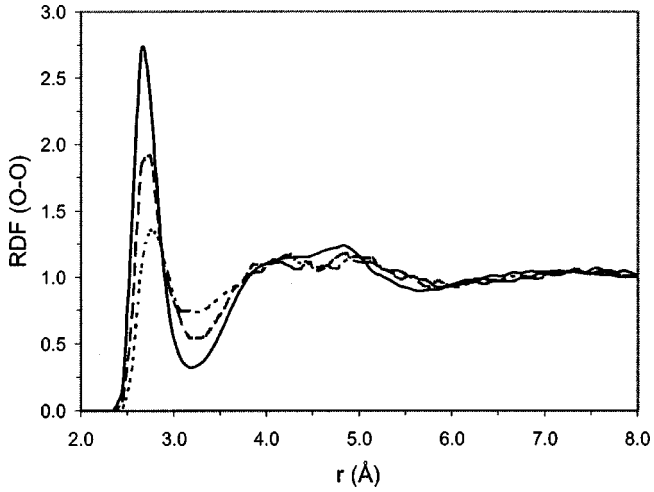

FIG. 6. O(formamide)–O(water) radial distribution function of methanol for the first (dotted line), second (dashed line), and the last (full line) cycles of the ASEP/MD procedure.

TABLE III. Optimization of the geometry of formamide. Cycle 0 is the *in vacuo* calculation. The average of the converged cycles and the difference between the *in vacuo* and *in solution* geometries are also shown. For comparison the results obtained in Ref. 8 are displayed as DFT/MM. Distances in angstrom, angles in degrees.

Distance or angle	ASEP/MD (50 ps)							
	0 (*in vacuo*)	1	2	3	4	5	10	15
N1–H2	1.017	1.022	1.025	1.028	1.030	1.031	1.030	1.029
N1–H3	1.020	1.025	1.029	1.029	1.030	1.032	1.032	1.033
N1–C4	1.359	1.343	1.335	1.334	1.331	1.329	1.331	1.329
C4–O5	1.229	1.240	1.248	1.248	1.253	1.254	1.252	1.254
C4–H6	1.114	1.109	1.107	1.106	1.106	1.105	1.106	1.105
H2–N1–H3	117.7	117.7	117.6	117.7	117.5	117.8	117.7	117.4
H2–N1–C4	122.6	121.9	121.5	121.6	121.1	120.6	121.0	120.6
N1–C4–O5	125.4	125.5	125.5	125.1	125.6	125.6	125.3	125.6
N1–C4–H6	112.1	113.0	113.5	113.8	113.7	113.8	113.9	113.8

Distance or angle	ASEP/MD				DFT/MM		
	50 ps		150 ps				
	Average	Variation	Average	Variation	*in vacuo*	*in solution*	Variation
N1–H2	1.029	0.013	1.030	0.013	1.016	1.040	0.024
N1–H3	1.033	0.013	1.033	0.013	1.018	1.041	0.023
N1–C4	1.329	−0.029	1.330	−0.029	1.366	1.338	−0.028
C4–O5	1.253	0.024	1.253	0.024	1.227	1.262	0.035
C4–H6	1.106	−0.008	1.106	−0.008	1.114	1.112	−0.002
H2–N1–H3	117.6	−0.1	117.5	−0.1	119.3	119.7	0.4
H2–N1–C4	120.8	−1.9	120.8	−1.9	121.7	119.8	−1.9
N1–C4–O5	125.5	0.2	125.6	0.2	124.4	124.6	0.2
N1–C4–H6	113.8	1.8	113.8	1.7	112.5	114.2	1.7

ing an increase of almost 60% with respect to the *in vacuo* value. The polarization of the solute is accompanied by a parallel increase in the structure of the solvent around the solute. This is clearly observed in the evolution of the oxygen–oxygen radial distribution function with the number of cycles of the self-consistent process. As polarization progresses, the height of the first peak increases and moves toward shorter distances. During this process, the solute charge distribution and the solvent structure become mutually equilibrated (Fig. 6).

The computed geometries in gas phase and solution are listed in Table III. The final *in solution* geometries are average values over the different simulations (we give the averages for 50 ps and 150 ps simulations). For comparison, we also include results obtained using the DFT/MD method developed at Nancy. The first conclusion is that the optimized geometry depends very little on the size of the simulation. The averages obtained with the 50 ps and 150 ps simulations are almost identical. Furthermore, despite the gradient fluctuations associated with the simulation size, the solute geometries are very stable: the standard deviations are <0.001 Å for the bond lengths and 0.2° for the bond angles. Although our values are not directly comparable with the results obtained by Chalmet and Ruiz-López[8] due to small differences in the basis set and the density functional, they are very similar qualitatively and quantitatively: the solvent induces a decrease in the CN distance, an increase in the CO distances, and negligible variations in the bond angles. These results can be interpreted simply by the stabilization of the zwitterionic electronic configuration by the electrostatic solute–solvent interactions, as was confirmed by the behavior of the net atomic charges obtained by a CHELP (Ref. 14) population analysis. As expected, the electronic density of the N atom decreases from gas phase to solution, whereas it increases on the oxygen atom. The strong polarization of the formamide molecule in aqueous solution is manifested in the change of the dipole moment from 4.1 D *in vacuo* to 6.6 D *in solution*.

Next, we studied the water molecule. In this case, in order to check the performance of the proposed method when different calculation levels are used, the calculations were performed at the MP2 level with an aug-cc-pVDZ basis

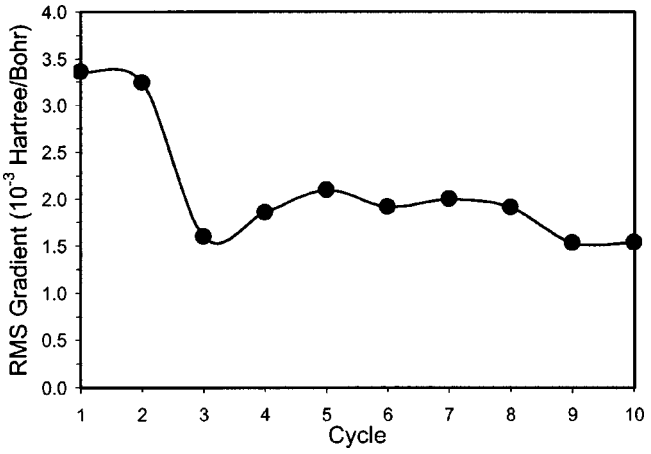

FIG. 7. Root mean square change in free-energy surface of water.

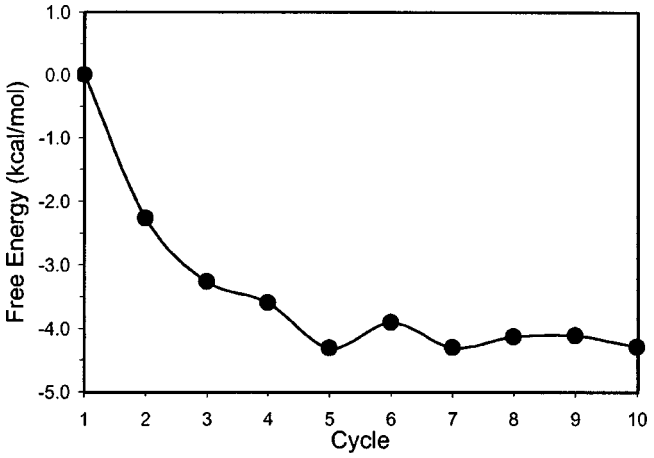

FIG. 8. Polarization free-energy change of a liquid water molecule during the optimization procedure.

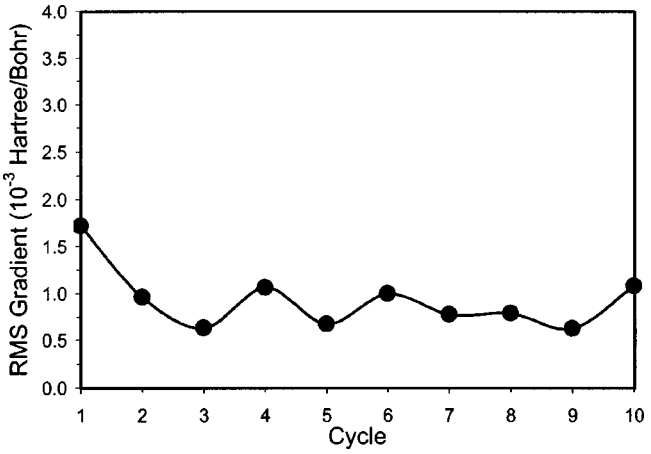

FIG. 10. Root mean square change in free-energy surface of methanol.

set.[17] As was the case for formamide, convergence is reached in about 4–5 cycles of the ASEP/MD procedure (see Fig. 7). The rms gradient value is also very similar, about 0.0015 hartree/bohr. The fluctuations in the free-energy, Fig. 8, are about 0.2 kcal/mol, clearly less than the fluctuations due to thermal effects. The water molecule undergoes a strong polarization during the solution procedure. The induced dipole moment, 0.77 D, compares well with the results obtained by other authors,[26] and with the estimated experimental[27] value, 0.70–1.0 D. The geometry changes are displayed in Fig. 9. The OH distance increases by 0.015 ± 0.002 Å. The bond angle does not change appreciably.

Finally, we studied liquid methanol. The basis set was of aug-cc-pVDZ (Ref. 17) quality. The calculations were performed at the DFT level. Convergence was reached in 3–5 cycles, Fig. 10,[13] and the rms gradient was somewhat lower in this case, 0.0008 hartree/bohr, than for formamide and water. The free-energy fluctuations, Fig. 11, are however, similar to those of water, about 0.3 kcal/mol. The *in solution* dipole moment was 2.30 D and the induced dipole moment 0.55 D. The methanol induced dipole moment is clearly lower than that obtained for water. This is so even though water's polarizability is almost half that of methanol. The explanation is to be found in the number of hydrogen bonds that each system forms. Thus, while methanol can form two hydrogen bonds, this number increases to four in the case of water. The changes in the geometry originated by the solvent are also less in methanol than in water. Only the OH bond is increased by about 0.008 Å. The rest of the molecule, Fig. 12, does not change appreciably. As was the case for the water and formamide, the fluctuations in the gradient have no appreciable effect on the geometry: the standard deviations are 0.001 Å for the bonds and 0.1°–0.4° for the angles.

The number of cycles necessary to reach convergence is very similar in the three systems considered, although in each case a different procedure was used to calculate the next solute geometry in the ASEP/MD cycle (calculation of the gradient, gradient and Hessian, or of the complete optimized geometry). Given that the computational times are very different depending on the method used, the best option is to use the simplest procedure, i.e., to calculate the new geometry by using only the gradient.

V. SUMMARY

In this article we have proposed a method for the optimization of molecules *in solution* that makes use of the mean field approximation and the free-energy gradient method. The method yields optimized stable geometries *in solution* in a very efficient way: the increase in computational time is minimal with respect to previous versions of the ASEP/MD that worked at a fixed geometry. As a new ingredient, we need only calculate the gradient, or the gradient and the Hessian, at each cycle of the ASEP/MD procedure. In general 5–6 cycles were enough to reach convergence. We have also

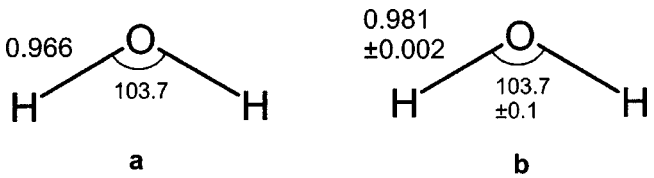

FIG. 9. Optimized geometries of water in the gas phase (a) and *in solution* (b).

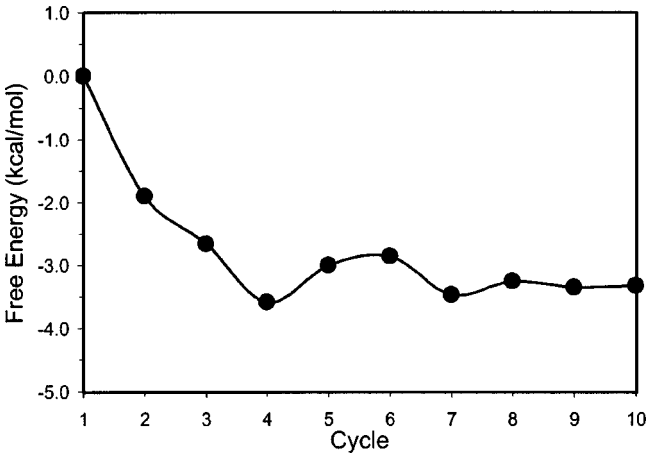

FIG. 11. Polarization free-energy change of a liquid methanol molecule during the optimization procedure.

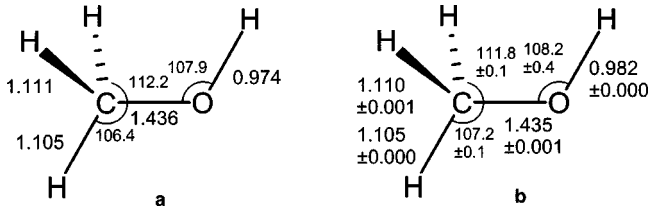

FIG. 12. Optimized geometries of methanol in the gas phase (a) and *in solution* (b).

shown that the fluctuation in the gradient and free-energy are related to the time of simulation. Except for formamide, where it was necessary to perform longer simulations (150 ps), for water and methanol simulations of 75 ps yielded to free-energy fluctuations that were lower than the thermal contribution. In the three molecules studied, and despite the gradient and free-energy fluctuations, the geometries obtained were very stable. The standard deviations were <0.001 Å for the bond length and $<0.2°$ for the bond angles.

It is interesting to note that, for the molecules considered, the solvent only induces small changes in the solute geometry. In fact, the influence that the basis set or the level of calculation has on the geometry is larger than the solvent influence. For instance, in the water system, the *in vacuo* OH bond distance varied by about 0.022 Å from the HF to the MP2 level, and by 0.031 from the HF to the DFT level. A similar behavior was presented by the other two systems. The solvent, however, induced large changes in the solute charge distributions, which became strongly polarized. Increases in the dipole moment between 30%–60% were found. At the same time, the polarization of the solute was accompanied by an increase in the structure of the solvent around the solute. The first peak of the radial distribution function shifted to shorter distances and its height increased by several units.

In the present ASEP/MD version there are two possible sources of error: the use of the mean field approximation and the different representations of the electrostatic solute charge distribution during the simulation (classical) and the optimization (quantum mechanical). In this and a previous paper, we have shown that the MFA only introduces small errors in the magnitudes evaluated (dipole moment, interaction energies, gradients, etc.) while it allows one to greatly reduce the number of quantum calculations to perform. The errors associated with the classical representation of the solute during the MD calculation can be reduced by improving the solute charge representation through the inclusion of additional charges on the solute molecule. We are working in this direction. In any case, these errors seem negligible when compared with the errors introduced by the rest of the approximations used in most QM/MM methods (neglect of intermolecular electron exchange, parametrization of the Lennard-Jones potential, etc.)

ACKNOWLEDGMENTS

This research was sponsored by the Dirección General de Investigación Científica y Técnica (BQU2000-0243) and by the Consejería de Educación y Juventud de la Junta de Extremadura (Project No. 2PR01A010).

[1] O. Tapia, in *Theoretical Treatment of Large Molecules and Their Interactions*, edited by Z. B. Maksic (Springer-Verlag, Berlin, 1991), Vol. 4, p. 435.
[2] M. L. Sánchez, M. E. Martín, I. Fdez. Galván, F. J. Olivares de Valle, and M. A. Aguilar, J. Phys. Chem. **106**, 4813 (2002).
[3] (a) J. Tomasi, R. Bonaccorsi, R. Cammi, and F. J. Olivares de Valle, J. Mol. Struct.: THEOCHEM **234**, 401 (1991); (b) J. Tomasi and M. Persico, Chem. Rev. **94**, 2027 (1994); (c) J. L. Rivail and D. Rinaldi, in *Computational Chemistry: Review of Current Trends*, edited by J. Leszczynski (World Scientific, Singapore, 1995); (d) C. J. Cramer and C. J. Truhlar, in *Reviews in Computational Chemistry*, edited by K. B. Kipkowitz and D. B. Boyd (VCH, New York, 1995), Vol. VI, p. 1.
[4] (a) S. Teno-no, F. Hirata, and S. Kato, Chem. Phys. Lett. **214**, 391 (1993); (b) S. Ten-no, F. Hirata, and S. Kato, J. Chem. Phys. **100**, 7443 (1994); (c) M. Kawata, S. Ten-no, S. Kato, and F. Hirata, Chem. Phys. **240**, 199 (1995); (d) M. Kawata, S. Ten-no, S. Kato, and F. Hirata, J. Phys. Chem. **100**, 1111 (1996); (e) M. Kinoshita, Y. Okamoto, and F. Hirata, J. Comput. Chem. **18**, 1320 (1997); (f) R. Akiyama and F. Hirata, J. Chem. Phys. **108**, 4904 (1998); (g) H. Sato, A. Kovalenco, and F. Hirata, *ibid.* **112**, 9463 (2000).
[5] (a) M. L. Sánchez, M. A. Aguilar, and F. J. Olivares del Valle, J. Comput. Chem. **18**, 313 (1997); (b) M. L. Sánchez, M. E. Martín, M. A. Aguilar, and F. J. Olivares del Valle, Chem. Phys. Lett. **310**, 195 (1999); (c) M. E. Martín, M. L. Sánchez, J. Olivares del Valle, and M. A. Aguilar, J. Chem. Phys. **113**, 6308 (2000); (d) M. E. Martín, M. L. Sánchez, M. A. Aguilar, and F. J. Olivares de Valle, J. Mol. Struct.: THEOCHEM **537**, 213 (2001); (e) M. L. Sánchez, M. E. Martín, M. A. Aguilar, and F. J. Olivares del Valle, J. Comput. Chem. **21**, 705 (2000); (f) M. E. Martín, M. L. Sánchez, F.J. Olivares del Valle, and M. A. Aguilar, J. Chem. Phys. **116**, 1613 (2002).
[6] (a) F. Hirata and P. J. Rossky, Chem. Phys. Lett. **83**, 329 (1981); (b) F. Hirata, P. J. Rossky, and B. M. Pettitt, J. Chem. Phys. **78**, 4133 (1983); (c) D. Chandler and H. C. Andersen, *ibid.* **57**, 1930 (1972).
[7] (a) M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, London, 1987); (b) J. A. McCammon and J. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, Cambridge, 1987).
[8] S. Chalmet and M. F. Ruiz-López, J. Chem. Phys. **11**, 1117 (1999).
[9] N. Okuyama-Yoshida, M. Nagaoka, and T. Yamabe, Int. J. Quantum Chem. **70**, 95 (1998).
[10] N. Okuyama-Yoshida, K. Kataoka, M. Nagaoka, and T. Yamabe, J. Chem. Phys. **113**, 3519 (2000).
[11] H. Hirao, Y. Nagae, and M. Nagaoka, Chem. Phys. Lett. **348**, 350 (2001).
[12] (a) A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976); (b) U. C. Singh and P. A. Kollman, J. Comput. Chem. **7**, 718 (1986); (c) M. J. Field, P. A. Bash, and M. Karplus, *ibid.* **11**, 700 (1990); (d) V. Luzhkov and A. Warshel, *ibid.* **13**, 199 (1992); (e) J. Ga, in *Reviews in Computational Chemistry*, edited by K. B Lipkowitz and D. B. Boyd (VCH, New York, 1996), Vol. 7, p. 119.
[13] (a) R. Cammi and J. Tomasi, J. Chem. Phys. **101**, 3888 (1994); (b) E. Cancès, B. Mennucci, and J. Tomasi, *ibid.* **109**, 260 (1998); (c) V. Barone, M. Cossi, and J. Tomasi, J. Comput. Chem. **19**, 404 (1998).
[14] L. E. Chirlian and M. M. Francl, J. Comput. Chem. **8**, 894 (1987).
[15] St. Amant and D. R. Salahub, Chem. Phys. Lett. **169**, 387 (1990).
[16] J. P. Perdew, Phys. Rev. B **33**, 8822 (1986).
[17] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[18] (a) T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989); (b) R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, *ibid.* **96**, 6796 (1992); (c) D. E. Woon and T. H. Dunning, *ibid.* **98**, 1358 (1993).
[19] GAUSSIAN 98, M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, Gaussian, Inc., Pittsburgh, Pennsylvania, 1998.
[20] K. Refson, MOLDY User's Manual Revision 2.10 (University of Oxford Press, Oxford, 1996) ftp.earth.ox.ac.uk/pub
[21] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[22] W. L. Jorgensen and C. J. Swenson, J. Am. Chem. Soc. **107**, 569 (1985).
[23] (a) W. L. Jorgensen, J. Phys. Chem. **90**, 1276 (1986); (b) L. Jorgensen, J. D. Madura, and C. J. Swenson, J. Am. Chem. Soc. **106**, 6638 (1984).
[24] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[25] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954); U. C. Singh, F. K. Brown, P. A. Bash, and P. A. Kollman, J. Am. Chem. Soc. **109**, 1607 (1987).
[26] (a) I. Tuñón, M. T. C. Martins-Costa, C. Millot, M. F. Ruiz-López, and

J.L. Rivail, J. Comput. Chem. **17**, 19 (1996); (b) P. Ahlström, A. Wallqvist, S. Engström, and B. Jönsson, Mol. Phys. **68**, 563 (1989); (c) K. Laasonen, M. Sprik, M. Parrinello, and R. Car, J. Chem. Phys. **99**, 9080 (1993); (d) G. Jansen, F. Colonna, and J. G. Angyán, Int. J. Quantum Chem. **58**, 251 (1996); (e) P. L. Silvestrelli and M. Parrinello, J. Chem. Phys. **111**, 3572 (1999).

[27] (a) D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water* (Oxford University Press, New York, 1969); (b) C. A. Coulson and D. Eisenberg, Proc. R. Soc. London, Ser. A **291**, 445 (1966); (c) E. Whalley, Chem. Phys. Lett. **53**, 449 (1978); (d) Y. S. Badyal, M. L. Saboungi, D. L. Price, S. D. Shasri, D. R. Haeffner, and A. K. Soper, J. Chem. Phys. **12**, 9206 (2000).