Incorporation of solvent effects into density functional calculations of molecular energies and geometries

Jan Andzelm, Christoph Kölmel, and Andreas Klamt

Additional information on J. Chem. Phys.
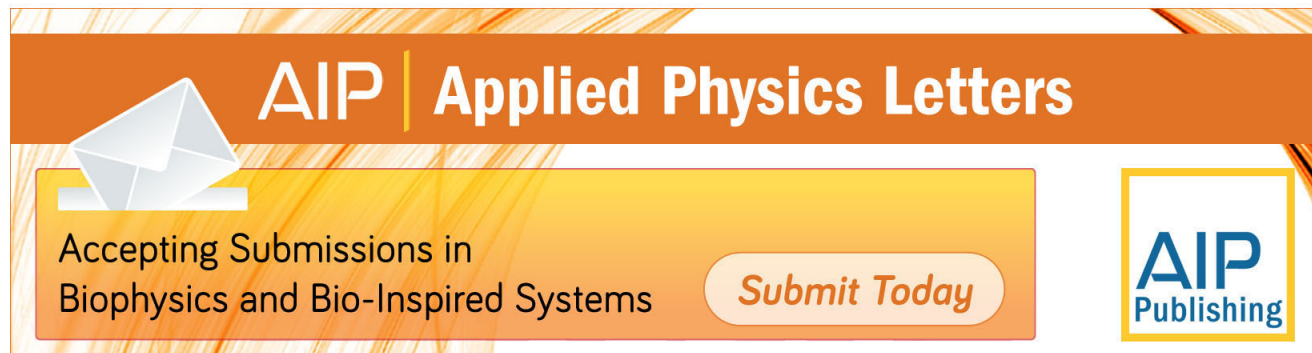
Incorporation of solvent effects into density functional calculations of molecular energies and geometries

Jan Andzelm and Christoph Kölmel

BIOSYM/Molecular Simulations, San Diego, California 92121

Andreas Klamt

Bayer AG, Q18, D-51368 Leverkusen-Bayerwerk, Germany

In this paper, we present the implementation of the ''conductorlike screening model'' (COSMO) into the density functional program DMol. The electronic structure and geometry of the solute are described by a density functional method (DFT). The solute is placed into a cavity which has the shape of the solute molecule. Outside of the cavity, the solvent is represented by a homogeneous dielectric medium. The electrostatic interaction between solute and solvent is modeled through cavity surface charges induced by the solvent. The COSMO theory, based on the screening in conductors, allows for the direct determination of the surface charges within the SCF procedure using only the electrostatic potentials. This represents the major computational advantage over many of other reaction field methods. Since the DMol/COSMO energy is fully variational, accurate gradients with respect to the solute coordinates can be calculated for the first time, without any restriction on the shape of the cavity. The solvation energies and optimized molecular structures are calculated for several polar solutes. In addition, the trends in basicity of amines and the relative stabilities of molecular conformers are studied. Our results suggest that for neutral solutes, agreement between calculated and experimental solvation energies of better than about 2 kcal/mol can be achieved. © *1995 American Institute of Physics.*

I. INTRODUCTION

Many biochemical and catalytic processes occur in solution where the chemical properties and processes are often very different from the gas phase. It is therefore very important to be able to simulate the effect of solvent on chemical reactions. The ''first principles'' quantum mechanical methods are capable of correctly describing the electronic and geometrical structures of solute molecules, including polarization effects, due to the presence of the solvent. However, in a quantum mechanical approach, only a limited number of solvent molecules can be considered explicitly due to the high cost of the calculations. The use of density functional methods (DFT) as implemented in DMol program allows to treat larger molecular systems than in the traditional, *ab initio* methods such as Hartree–Fock (HF) or Møeller–Plesset perturbation theory (MP2). However, even with DFT a simplified model of the solvent has to be used. The most successful model so far is the polarizable continuum model (PCM) by Tomasi and co-workers.[1,2]

In the original PCM approach, the solute interacts with the solvent represented by a dielectric continuum model. The solute molecule is embedded into a cavity surrounded by a dielectric continuum of permittivity ϵ. The dielectric continuum is polarized by the charge distribution of the solute which results in a charge distribution on the cavity surface. The original PCM approach requires a coupled iterative procedure. For a given charge distribution within the cavity, the surface screening charges are calculated by solving the Poisson equation. The calculated screening charges are then used as a perturbation included in the Hamiltonian for electronic structure calculations of the solute. The new, perturbed density of the solute is used again for Poisson calculations and

the procedure is repeated until self-consistency is achieved. There are several applications of this model for calculating the solvation energy for cavities of general shape cavity at the *ab initio*,[1–6] DFT,[6–11] and semiempirical[12,15] level. In the case of an ellipsoidal cavity, the calculation of analytical gradients was accomplished as well.[15,10] The theory of analytical gradients was presented recently.[16,17] The values of the solvation energies were found to be very sensitive to the size of the cavity. Using optimized van der Waals radii[2,5,8,18] leads to quite satisfactory estimates for solvation energies.

Recently, Klamt *et al.*[15] proposed a new method, called the ''conductorlike screening model'' (COSMO), based on the screening in conductors which is noniterative and allows for the calculation of accurate gradients without cavity shape constraints. The surface charges are calculated directly from the electrostatic potential of the charge distribution within the cavity. This follows from imposing the boundary condition of vanishing electrostatic potential on the surface of a conductor. The related, noniterative approaches by Hoshi *et al.*[19] and Coitino *et al.*[20] require evaluation of electric fields on the cavity surface instead of electrostatic potentials.[15] The COSMO approach leads to better numerical accuracy, and is simpler, particularly for the analytic gradient evaluation. The electrostatic part of the solvation energy is a quadratic function of the solute charge distribution and it can be included in the solute Hamiltonian. The self-consistent process of optimizing the solute charge density is done with the screening potential simultaneously included in the SCF cycle. Therefore, the COSMO method is fully variational and accurate gradients can be readily calculated. The COSMO method has been implemented already at the semiempirical

level.[15] Recently, COSMO energy calculations were done within *ab initio* and DFT methods.[6]

In this paper, we present the theory of COSMO energy and gradient implementation within the density functional program—DMol. We selected DMol because this is a well-developed and validated methodology which allows for rapid and accurate DFT calculations for molecular systems.[21,22] Furthermore, DMol is unique because of its high quality numerical basis sets which exhibit a small BSSE error and it also provides diffuse orbitals, which are necessary to describe polarization of the electron density. The total electron density is approximated using a projection rather than a fitting technique which improves both performance and accuracy of the method.

The new procedure was tested in calculations of solvation energies and geometries for a small number of neutral and charged molecules. The DMol/COSMO method provides results of accuracy comparable to other calculations,[1–10,16–20] however it is faster and geometry optimization can be done for solute cavities of any shape.

II. THEORY

Within the COSMO[15] approach, the total electrostatic energy of the system comprising of solute and the solvent as a dielectric continuum can be written in matrix form as follows:

$$E(q) = 1/2\, QCQ + QBq + 1/2\, qAq, \tag{1}$$

where q is the vector of the screening charges on the surface of the cavity and $Q = \rho + Z$ denotes all the solute charges such as electron density $\rho(r)$ and nuclear charges, Z. C is the standard Coulomb matrix, B describes the electrostatic interaction of solute and screening unit charges, and A is the electrostatic interaction of screening unit charges. Thus the first term in Eq. (1) is the standard, gas phase, Coulomb energy, and the remaining terms are due to the solvent effects. The complete definition of the B, A matrices as well as the design of cavity and distribution of screening charges are the same as in Ref. 15.

The optimal set of surface screening charges q results from minimizing the total electrostatic energy (1), with respect to q. This procedure yields

$$q^* = -A^{-1}BQ. \tag{2}$$

This equation can also be deduced from imposing the boundary condition of vanishing electrostatic potential on the cavity surface. The total electrostatic energy of the system, with the optimal screening charges q^* now becomes

$$E(q^*) = 1/2\, QCQ - 1/2\, QDQ; \quad D = BA^{-1}B, \tag{3}$$

where the second term is the screening energy and D is the dielectric operator matrix based on the COSMO approach.[15]

The functional form of the screening energy, Eq. (3), indicates that the implementation of COSMO into a first principle program formally requires evaluation of the two-electron interactions in full analogy with Coulomb repulsive interactions. Fortunately, no two electron integrals have to be calculated, since the screening energy, ΔE can be rewritten as

$$\Delta E = -1/2\, QDQ = -1/2\, QBq^*. \tag{4}$$

Thus the ΔE is the interaction of the solute charge distribution, Q, with the screening charges q^*. This term requires the calculation of only one-electron integrals as was recently implemented.[15,6]

Before we present the theory of the COSMO implementation in DMol,[21,22] we rewrite the total energy of DMol as follows:

$$E_{\text{DMol}} = E_{\text{kin}} + E_{\text{xcor}} + E_{\text{elst}}, \tag{5}$$

where E_{kin}, E_{xcor}, and E_{elst} stand for kinetic, exchange-correlation, and electrostatic energy, respectively. The DMol electrostatic energy is given as

$$E_{\text{elst}} = 1/2\langle Z|C|Z\rangle + \langle \rho|C|Z\rangle + \langle \rho|C|\rho_-\rangle - 1/2\langle \rho_-|C|\rho_-\rangle, \tag{6}$$

where $\langle Z|C|Z\rangle$ is the internuclear repulsion energy, $\langle \rho|C|Z\rangle$ describes the nuclear–electron attraction energy due to the interaction between the nuclear charges, Z and electron density, ρ. The last two terms of Eq. (6) describe the electron–electron repulsion, ρ_- is the projected (or input) density. The Coulomb electron–electron interaction energy uses ρ_- to facilitate calculations and it differs from the true energy by terms of second and higher order in $(\rho - \rho_-)$.[23] The multipolar representation of ρ_- allows for the efficient solution of the Poisson equations yielding the electrostatic potential.[21]

According to Eq. (3), the COSMO electrostatic energy is of analogous form to the DMol electrostatic energy, with the Coulomb operator C replaced by the dielectric operator, D. The COSMO energy contribution to the DMol energy can therefore be written as

$$E_{\text{elst}}^{\text{COSMO}} = -1/2\langle Z|D|Z\rangle - \langle \rho|D|Z\rangle - \langle \rho|D|\rho_-\rangle + 1/2\langle \rho_-|D\rho_-\rangle. \tag{7}$$

The total energy of the solute in the presence of the electrostatic field from the solvent [Eqs. (5) and (7)] is now minimized by applying the usual Kohn–Sham approach.[24,25] In this method, electron density, ρ is expressed as a sum of the square of molecular orbitals, ϕ, weighted by occupation numbers n_i

$$\rho = \sum_i n_i \phi_i^2. \tag{8}$$

By assuming that the projected density (ρ_-) does not depend on the molecular orbitals, the optimal molecular orbitals which minimize the total energy are a solution of the Kohn–Sham equations

$$F\phi_i = e_i\phi_i,$$
$$F = T + V_{\text{xc}} + V_{\text{elst}} + V_{\text{elst}}^{\text{COSMO}}, \tag{9}$$

where T, V_{exc} are the standard kinetic and exchange correlation operators and V_{elst}, $V_{\text{elst}}^{\text{COSMO}}$ define the electrostatic potentials of DMol and COSMO, respectively

$$V_{\text{elst}} = C|Z + \rho_-\rangle,$$
$$V_{\text{elst}}^{\text{COSMO}} = -D|Z + \rho_-\rangle = -DQ_- = Bq_-^*. \tag{10}$$

Equation (10) was simplified by using the definition of the dielectric operator D and surface screening charges [Eq. (2)–(4)]. The electrostatic potentials are generated by solute and by the screening charges.

This completes the SCF procedure for the DMol/COSMO total energy calculation. At first, using a projected (input) electron density ρ_- the screening charges, q_-^* are calculated using Eq. (2). Next, the COSMO electrostatic potential is computed according to formula 10. This potential is present in every SCF cycle. It is worthwhile to emphasize

again that this direct incorporation of the solvent effects within the SCF procedure is a major advantage of the COSMO scheme.

The first derivatives of the total DMol/COSMO energy with respect to nuclear coordinates are straightforward to evaluate.

It is useful to apply the definition of the dielectric operator, D [Eq. (3)] while differentiating $E_{\text{elst}}^{\text{COSMO}}$ [Eq. (7)] with respect to an arbitrary coordinate, x of some solute atom. After some rearranging we can write

$$
\begin{aligned}
\frac{dE_{\text{elst}}^{\text{COSMO}}}{dx} &= \frac{d}{dx}\left\{ -\langle \rho + Z | BA^{-1}B | \rho_- + Z \rangle + \frac{1}{2}\langle \rho_- + Z | BA^{-1}B | \rho_- + Z \rangle \right\} \\
&= -\left\langle \frac{d}{dx}[(\rho+Z)B]\middle| A^{-1} \middle| B(\rho_-+Z) \right\rangle - \left\langle (\rho+Z)B \middle| A^{-1} \middle| \frac{d}{dx}[B(\rho_-+Z)] \right\rangle + \left\langle \frac{d}{dx}[(\rho_-+Z)B] \middle| A^{-1} \middle| B(\rho_-+Z) \right\rangle \\
&\quad - \left\langle (\rho+Z)B \middle| \frac{dA^{-1}}{dx} \middle| B(\rho_-+Z) \right\rangle + \frac{1}{2}\left\langle (\rho_-+Z)B \middle| \frac{dA^{-1}}{dx} \middle| B(\rho_-+Z) \right\rangle.
\end{aligned}
\tag{11}
$$

Using the definition of the screening charges [Eq. (2)] and the relationships

$$
\frac{dA^{-1}}{dx} = -A^{-1}\frac{dA}{dx}A^{-1}
$$

and

$$
\frac{d}{dx}[(\rho+Z)B] = \frac{d\rho}{dx}B + \frac{dZ}{dx}B + (\rho+Z)\frac{dB}{dx},
$$

we can rewrite Eq. (11) as follows:

$$
\begin{aligned}
\frac{dE_{\text{elst}}^{\text{COSMO}}}{dx} &= -\left\langle \frac{d\rho}{dx}Bq_-^* \right\rangle + \left\langle \rho\frac{dB}{dx}q_-^* \right\rangle \\
&\quad + \left\langle q^*\left\{ B\frac{dZ}{dx} + \frac{dB}{dx}Z \right\} \right\rangle \\
&\quad + \left\langle (q^*-q_-^*)\left\{ B\frac{d\rho_-}{dx} + \frac{dB}{dx}\rho_- \right\} \right\rangle \\
&\quad - \frac{1}{2}\left\langle q^*\frac{dA}{dx}q^* \right\rangle + \left\langle q^*\frac{dA}{dx}q_-^* \right\rangle.
\end{aligned}
\tag{12}
$$

Formula (12) can be easily interpreted. The first term is the Pulay–COSMO term, and arises from the incomplete molecular basis set. This term is calculated in exactly the same way as the original Pulay contribution in DMol,[21] because the COSMO screening potential, Bq_-^* [see Eq. (10)] is already included in the Kohn–Sham operator.

Here, dB/dx is the derivative of the B matrix with respect to the position of the screening charges q. This gradient contribution is zero if coordinate x and screening charge belong to the same atom. The dZ/dx derivative is a derivative of the product ZB with respect to the position of the atom with coordinate, x.

One can notice a symmetry in formula (12). The first term generates a force due to charge density ρ, toward screening charges q_-^*, whereas the second term is the response of the screening charges (cavity) to that force. Similarly the third term represents a force between nuclear charges, Z and the cavity and the opposite force of cavity towards nuclei. The fourth term vanishes in the case of $q^* = q_-^*$, i.e., a perfect projection to the electron density. This term is specific to the DMol method, which uses the projected density ρ_- to evaluate the electrostatic potential [Eq. (10)]. The last two terms are the derivatives of the A matrix. In the case of $q^* = q_-^*$ these two terms become equivalent to the last term of the Eq. (13) in the original derivation of COSMO forces.[15]

Calculation of integrals in Eq. (12) is done using the numerical integration technique in DMol. This introduces an additional inaccuracy for the gradient terms including the dB/dx operator. This is due to the fact that the singularity at the position of the screening charges is not treated properly, since there is no numerical integration grid centered at the screening charges. To remedy this accuracy problem, we use the translational invariance condition for the gradients

$$
\sum \frac{d}{dx}(\rho Bq^*) = 0,
\tag{13}
$$

where summation is done over all atoms of the solute. This condition implies that the calculation of $\langle \rho\, dB/dx\, q^* \rangle$ terms can be avoided and instead the $\langle d\rho/dx\, Bq^* \rangle$ terms have to be evaluated. The gradient of the COSMO potential is then calculated with the same accuracy as all other contributions to the DMol gradient. Further details of the gradient implementation and optimization of the algorithm will be published elsewhere.[26]

It is well known that the dielectric screening energies for a given geometry scale as $(\epsilon-1)/(\epsilon+x)$, with the dielectric permittivity ϵ of the screening medium, where x is in the range 0–2. (For detailed discussion see Ref. 15.) In the present paper we adopted a value of $x=0.5$, following recommendation in Ref. 15. Therefore, the total electrostatic COSMO contribution to energy [Eq. (7)], the COSMO potential [Eq. (10)] and COSMO gradients [Eq. (12)] are scaled by the factor of

$$f(\epsilon)=(\epsilon-1)/(\epsilon+0.5).$$

This causes a small error of $1/(2\epsilon)$ which is negligibly small for polar solvents such as water which we study in the present paper.

III. COMPUTATIONAL DETAILS

All DFT calculations were performed within the Kohn–Sham formalism[23,24] using the DMol program.[21,22,27] High quality double numerical atomic basis sets including polarization functions (DNP) were used. The fine grid and standard DMol partitioning scheme were employed, which amounts to about 2000 integration points per atom. In the present work, two DFT Hamiltonians were used: local spin density (LSD) Hamiltonian and nonlocal (NLSD), gradient corrected approach. The LSD method employs the Vosko, Wilk, and Nusair (VWN)[28] parametrization of the Ceperley–Alder Monte Carlo results for a homogeneous electron gas. The NLSD method, denoted as BP in the present paper, is a combination of the Becke[29] gradient corrected exchange energy and the Perdew and Wang[30] gradient corrected correlation energy. Local parametrization due to Perdew and Wang was used for the BP calculations. Fully self-consistent calculations were performed for both LSD and NLSD potentials. These DMol parameters lead to quite accurate results for molecular structures and energetics as has been shown elsewhere.[22,31] Optimization of the molecular structures was done by using analytic gradients and a Hessian update method as recently described by Baker.[32] Optimization was done until change in a value of maximum gradients was less

than 0.001 atomic units. A tight SCF convergence criterion $(10^{-7}$ a.u.) was used for the SCF calculation.

The calculation of solvation effects was done within our DMol/COSMO program.[27] The details of cavity construction were already reported by one of us[15] and will be explained in more details elsewhere.[26] A dielectric permittivity of $\epsilon=78.4$ was used which corresponds to water at room temperature. Since the solvation energy is very sensitive to the size of the cavity, we investigated several commonly used sets of atomic radii. According to Tomasi and co-workers,[2] atomic radii larger by 20% than the standard van der Waals radii perform best in continuum models. Basch et al. suggest to increase the van der Waals radii by 25% and to use different radii for polar and apolar hydrogens.[18] The hydrogen radius was optimized in a recent study by Tannor et al.[5] Since detailed optimization of the surface cavity parameters is beyond the scope of this paper, we adopted the atomic radii by Bondi,[33] which are known to provide good results for organic molecules. These parameters have been tested extensively in calculation of solvation energies.[8] Following Chen et al.[8] we use a smaller value for the oxygen radius.

It is clear that we can further improve the accuracy of the present approach by performing an optimization of the atomic radii which is currently under investigation in our lab.

The present DMol/COSMO approach evaluates only the electrostatic contribution to the solvation energy. The total solvation free energy includes the nonelectrostatic terms which account for formation of the cavity and dispersion effects and are generally positive.[2] These terms can be expressed as a linear function of the cavity surface. The linear coefficients are usually obtained from fitting the free energies of hydration for linear-chain alkanes surface area.[2,34] In the present paper we adopted the procedure by Chen et al.[8] and used their values of nonelectrostatic terms to "correct" experimental values of free solvation energies. The calculated DMol/COSMO results of electrostatic solvation energies can be directly compared with experimental energies "corrected" for the nonelectrostatic terms.

TABLE I. Calculated and experimental solute dipole moments (μ, in Debye) and free energies of hydration (ΔG, in kcal/mol) for a set of test molecules. Experimental data of the gas phase dipole moments from Ref. 36. The calculated dipole moments in solution are in parentheses. Results with local, VWN and nonlocal, BP DFT Hamiltonians are given. Experimental values of hydration energies (ΔG_{exp}) for neutral and ionic species are taken from Refs. 34 and 37, respectively. The experimental energies corrected for nonelectrostatic terms (ΔG_{expc}) are taken from (a) Ref. 8 (b) Ref. 5 (c) Ref. 6.

Molecule	ΔG_{exp}	ΔG_{expc}	ΔG_{VWN}	ΔG_{BP}	μ_{exp}	μ_{VWN}	μ_{BP}
H_2O	−6.3	−8.7[a]	−9.3	−8.7	1.85	1.86(2.41)	1.80(2.33)
NH_3	−4.3	⋯	−7.3	−6.9	1.47	1.62(2.32)	1.58(2.25)
$HCOOH$	−5.5	⋯	−9.0	−8.6	1.41	1.48(2.31)	1.45(2.24)
CH_3OH	−5.1	−6.9[a]	−6.6	−6.3	1.70	1.59(2.18)	1.59(2.19)
$HCONH_2$	⋯	⋯	−13.2	−12.8	3.73	3.87(5.83)	3.81(5.74)
CH_3NH_2	−4.6	−6.4[a]	−5.9	−5.7	1.31	1.32(2.01)	1.35(1.99)
$(CH_3)_2NH$	−4.3	−6.4[b]	−4.7	−4.4	⋯	0.89(1.48)	0.95(1.52)
$(CH_3)_3N$	−3.2	−5.4[b]	−3.4	−3.0	⋯	0.35(0.82)	0.44(0.91)
NH_4^+	−77	−78.7[c]	−86.5	−86.5			
$CH_3NH_3^+$	−68	−71.8[a]	−76.7	−76.2			
$(CH_3)_2NH_2^+$	−61	−63.1[c]	−68.8	−68.0			
$(CH_3)_3NH^+$	−57	−59.2[c]	−62.7	−61.8			
NO_2^-	−70	−71.8[c]	−72.8	−73.0			
$CH_3CO_2^-$	−75	−79.3[a]	−73.7	−73.8			

TABLE II. Relative proton affinities of nitrogen bases (in kcal/mol).

Nitrogen base	Gas phase			Solvent		
	Expt.[a]	VWN	BP	Expt.[b]	VWN	BP
NH_3	0	0	0	0	0	0
CH_3NH_2	9.0	10.6	11.6	1.9	2.0	2.4
$(CH_3)_2NH$	15.3	16.8	18.8	2.1	1.7	2.9
$(CH_3)_3N$	19.8	20.2	23.2	0.8	0.4	2.3

[a]Reference 38.
[b]Calculated using formula (15) with data from Ref. 36.

IV. RESULTS AND DISCUSSION

A. Hydration energies

In Table I, the results of DMol/COSMO calculations are reported for our test molecules. Dipole moments of polar molecules and hydration energies including cations and anions are given at VWN and BP levels of theory. A full geometry optimization was performed in all cases.

The calculated hydration energies for polar, neutral molecules are within ~2 kcal/mol of the available experimental data corrected for nonelectrostatic contributions. In the case of H_2O the BP potential seems to be superior to the local VWN potential, however this is not the case for the methylamines. Gas phase dipole moments calculated are in very good agreement with experimental data at both VWN and BP levels of theory. This is due to the proper asymptotic behavior of the basis functions employed in DMol and has already been observed before.[35] Dipole moments in the presence of solvent are consistently larger than in gas phase and overall similar at both levels of density functional theory. The calculated dipole moment of water in aqueous solution of ~2.4 Debye is in very good agreement with experimental estimates. The results of hydration energies for solute ions are considerably less accurate than for the neutral molecules. The same trend was observed for other continuum electrostatic models.[6,9] For cations the present method consistently overestimates the hydration energies, however the relative values of ΔG for methyl substituted amines are in reasonable agreement with experimental data (see next section).

B. Relative basicities

Proton-transfer processes have been extensively studied with various quantum chemistry techniques.[38] The measure of the acid and base strengths of the gas-phase molecule can be the absolute gas-phase basicity or the absolute proton affinity. The latter is defined as the negative enthalpy of protonation process

$$A + H^+ \rightarrow AH^+ \tag{14}$$

and includes the changes in the electronic energy, zero point energy, the temperature dependent portion of vibrational energy, translational energy upon reaction. A knowledge of the absolute proton affinity is frequently not needed and the relative strength of different bases may be calculated instead. In the present paper the relative basicity of methylated amines was determined in gas phase and in aqueous solution. This simplifies our calculations since there is no need to evaluate

translational energies and we can neglect the fairly small temperature dependent portion of vibrational energy. The relative changes of zero-point vibration energies for the methylated amines in gas phase are quite constant (about 8 kcal/mol),[38] and we therefore did not incorporate this effect for our gas phase and solvent study.

Experimental values of pKa of the methylated amines are known in aqueous solution.[36] Using the formula

$$\Delta G = RT2.303pKa$$
$$= 1.364pKa \quad (T = 298.13 \text{ K}), \tag{15}$$

we can obtain experimental changes in free energy of the reaction (14). These results can be related to the proton affinities calculated in gas phase, PA, by using a thermochemical cycle as follows:

$$\Delta G = PA - E_s(AH^+) + E_s(A) + E_s(H^+). \tag{16}$$

Here, E_s is the electrostatic solvation energy. The hydration energies of AH^+ and A were calculated using the DMol/COSMO model, whereas $E_s(H^+)$ was taken form experimental data.[39] There is an experimental uncertainty of about 3 kcal/mol for $E_s(H^+)$;[39,8] however, since we report only relative values of ΔG, this does not affect our conclusions. Similarly, we do not need to consider the translational entropy of the proton in formula (16).

Table II includes the relative gas phase proton affinities and the relative free energies of solvation for methylated amines, calculated using DMol/COSMO model. Two Hamiltonians, local, VWN, and nonlocal, BP were used in the calculations.

Examination of Table II shows that the VWN model reproduces trends in gas phase proton affinities within 2 kcal/mol, while the BP approach is somewhat less accurate. The
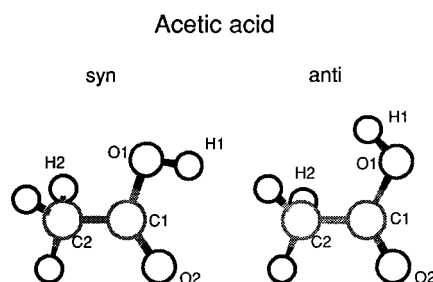

Acetic acid

FIG. 1. Acetic acid conformers.

TABLE III. Geometrical parameters in gas and solvent (sol) phase for *syn*- and *anti*-forms of acetic acid. Bond distances in Å, bond angles in degrees.

| Parameter | Syn-form | | | | Anti-form | | | |
| | VWN | | BP | | VWN | | BP | |
	gas	sol	gas	sol	gas	sol	gas	sol
C1–C2	1.482	1.475	1.505	1.498	1.492	1.479	1.515	1.503
C1–O1	1.354	1.342	1.377	1.364	1.360	1.345	1.385	1.369
C1=O2	1.216	1.227	1.224	1.235	1.209	1.224	1.217	1.231
O1–H1	0.995	0.999	0.990	0.996	0.988	0.993	0.985	0.988
C2–H2	1.104	1.105	1.100	1.101	1.106	1.106	1.102	1.101
H1O1C1	104.6	107.0	104.8	106.9	108.9	109.8	108.6	109.3
C2C1O1	111.9	112.8	111.7	112.6	115.0	116.1	115.3	116.1
O1C1O2	121.9	121.5	122.2	121.8	119.9	118.5	119.5	118.3

DMol/COSMO calculations reproduce the experimental trends quite well. The BP Hamiltonian seems to be slightly more accurate, by identifying the $(CH_3)_2NH$ molecule as the strongest base. However, the relative BP-calculated basicity of $(CH_3)_3N$ molecule is too large by 1.5 kcal/mol, in comparison with experimental values. Both Hamiltonians predict that the NH_3 molecule is the weakest base of the nitrogen bases studied.

The results of DMol/COSMO calculations confirm the importance of the solvent on the basicity of the methylated amines. In gas phase, the basicity increases with the number of methyl groups while in the aqueous environment all the methylated amines studied in this paper are of comparable basicity.

C. Relative stabilities of molecular conformers

In the following, the relative stability of the *syn*- and *anti*-forms of acetic acid (as displayed in Fig. 1) are studied. Full geometry optimization was performed using the analytic gradient code of the DMol/COSMO program. Table III displays geometry parameters of the conformers for gas and solvent phase. The VWN and BP Hamiltonians were employed in the calculations. Examination of the results in Table III shows that the changes in geometry upon solvation do not exceed 0.02 Å for bond distances and about 3 deg for bond angles. The more covalent bonds, such as C1C2 and C2H2 are affected less than the more ionic bonds involving oxygen atoms. The trends in geometry changes are fairly consistent for both conformers and Hamiltonians. The single bond C1O1 is contracted while the double bond C1O2 expands upon solvation. This effect is due to the larger partial

charge of the doubly bonded oxygen atom in solution. The H1O1C1 angle of the *syn*-form and O1C1O2 of the *anti*-form seem to be most sensitive to solvation effects. The differences between VWN and BP potentials are most visible for bond distances and they are of the same magnitude as changes due to solvation effects.

With exception of the angle H1O1C1 of the *syn*-form we may conclude that the *anti*-form undergoes the more significant change upon solvation. This conclusion is supported by the considerable change of the dipole moment and the gain in solvation energy during the geometry optimization. Table IV presents results of solvation energies and dipole moments for the conformers studied here.

The *syn*-form is the most stable conformer in gas phase by 5.1 and 4.8 kcal/mol for VWN and BP potential, respectively. This is in a good agreement with MP3[40] and other BP[8] calculations. In aqueous solution the *anti*-form is much more stabilized than the *syn*-form. The solvation energy is larger by about 3.4 kcal/mol with the VWN potential compared to the BP potential. As a result the *anti*-form is only about 1.7 kcal/mol less stable than the *syn*-form in very good agreement with the experimental estimate of 1–2 kcal/mol[41,42] and recent solvation studies.[8]

The main reason for that seems to be a significantly larger dipole moment of the *anti*-form which leads to stronger interaction with water molecules of the solvent. The dipole moment of the *syn*-form increases by 0.9 Debye, however the *anti*-form is more polar by about 2 Debye due to solvation effects. It is interesting to note that the gain in solvation energies following geometry optimization is relatively modest, i.e., only 0.5 kcal/mol. Thus the reorganiza-

TABLE IV. The solvation energies (E_s) (in kcal/mol) and dipole moments μ (in Debye) of *syn*- and *anti*-forms of acetic acid in gas phase and in aqueous solution. Relative energies of *anti*-form with respect to *syn*-form are given (RE).

| Property | Syn-form | | | | Anti-form | | | |
| | VWN | | BP | | VWN | | BP | |
	gas	sol	gas	sol	gas	sol	gas	sol
μ	1.749	2.630	1.674	2.518	4.335	6.193	4.192	5.974
E_s[a]	9.11	9.42	8.60	8.86	12.36	12.70	11.60	11.94
RE	0	0	0	0	5.10	1.79	4.77	1.70

[a]The entry marked gas is a solvation energy calculated at gas phase geometry, whereas sol means energy obtained for fully optimized structure.

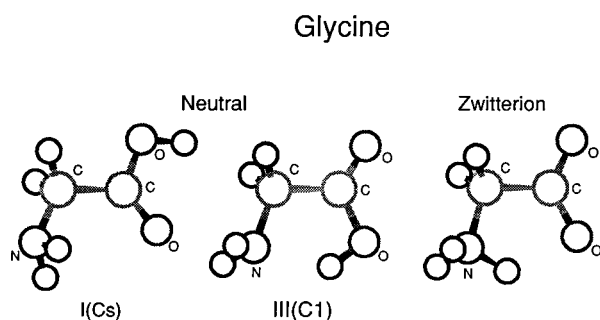## Glycine

Neutral   Zwitterion

I(Cs)   III(C1)

FIG. 2. Glycine conformers examined in this study.

tion of the electron distribution upon solvation is mainly responsible for the solvation effect.

D. Structure of glycine in aqueous environment

Conformations of glycine molecule have been studied extensively using various theoretical methods in both the gas phase and the solvent.[15,43–45] Although glycine is known to exist as a neutral form in the gas phase, in the condensed phase the zwitterion form becomes significantly more stable. In the present paper two of the most stable conformers of glycine in gas phase were considered. These are the structures denoted as I and III in the recent paper by Barone *et al.*[43] Microwave studies[46] indicate that structure I is more stable by about $1.4+/-0.4$ kcal/mol than structure III. However, both the VWN and BP calculations predict that structure III is more stable by 2.5 and 0.3 kcal/mol, respectively. These results can be compared with results by Barone *et al.*[43] who report the stabilization of structure III by 3.9 and 0.6 kcal/mol with VWN and BLYP Hamiltonians, respectively. The most accurate hybrid HF/DFT method predicts structure I to be more stable than structure III by only 0.2 kcal/mol.[43]

Starting from the neutral conformation (structure III), optimization in aqueous solution as simulated by the DMol/COSMO model leads to the zwitterion structure (Z) (see Fig. 2). In the zwitterion the proton from the carboxyl group is transferred to the amino group, while the structure of the amino-acid backbone is not changed significantly. One can notice (see Table V) that the N–C, C–C, and C=O bonds are elongated in the zwitterion, while the single bond C–O contracts by about 0.06 and 0.08 Å for VWN and BP calculations, respectively. It is well known[47] that the local spin density approach (VWN) leads to overestimation of the stability of hydrogen bonds and that hydrogen bonds are too short. The gradient corrected DFT approach (BP) significantly improves description of the hydrogen bonds.[47,43] The hydrogen bonds O–H and N–H (see Table V) are clearly too short at the VWN level and the BP calculations lead to an elongation of the hydrogen bonds.

Transfer of the proton bends the H atoms of the amino group away, while the single bonded oxygen atom follows the proton and bends towards amino group. The hydrogen bond in the neutral conformer III, between N and H, is replaced by the weaker hydrogen bond between O and H in the zwitterion. Similar qualitative results were reported from AM1/COSMO calculations.[15] However, the magnitude of this effect is completely different.

The AM1 optimization leads to the stable zwitterion structure in the gas phase which has an energy 42 kcal/mol higher than the neutral species. The DMol optimization does not find the zwitterion structure in the gas phase. The dielectric screening for the gas phase geometry at the AM1/COSMO level stabilizes the zwitterion form by 43 kcal/mol.[15] Relaxation of the geometry which results in significant weakening of the O–H hydrogen bond stabilizes the zwitterion structure by additional 7.3 kcal/mol. The DMol/COSMO calculations lead directly from the neutral structure III to the zwitterion. We were unable to find a stable conformation III in the aqueous solution.

With the gas phase geometry of structure III in DMol/COSMO energy calculations we find a solvation energy of 16.0 and 14.6 kcal/mol for the VWN and BP Hamiltonians, respectively. DMol/COSMO geometry optimization leads to the zwitterion with an additional gain of 5.8 and 6.8 kcal/mol for VWN and BP calculations, respectively. In summary, both the screening of the solvent and geometry relaxation stabilize the zwitterion by about 21 kcal/mol in comparison to the neutral, gas phase structure III.

TABLE V. Geometrical parameters in gas and solvent (sol) phase for neutral (I,III) and zwitterion conformers of glycine. Bond distances in Å, bond angles in degrees.

| Parameter | Conformer III[a] | | | | Conformer I | | | |
| | VWN | | BP | | VWN | | BP | |
	gas	sol	gas	sol	gas	sol	gas	sol
O–H	1.031	1.692	1.013	1.879	0.994	0.997	0.989	0.992
N–H	1.724	1.080	1.830	1.052	1.031	1.033	1.028	1.029
N–C	1.459	1.472	1.481	1.502	1.436	1.441	1.461	1.464
C–C	1.515	1.518	1.536	1.541	1.495	1.490	1.523	1.516
C–O	1.331	1.272	1.356	1.279	1.348	1.336	1.374	1.360
C=O	1.214	1.252	1.221	1.266	1.216	1.224	1.223	1.230
HNC	112.6	100.2	112.1	103.4	108.6	108.5	108.3	108.5
NCC	110.3	106.8	111.0	108.3	115.4	115.0	116.0	115.8
CC–O	112.2	114.9	113.1	116.3	112.3	112.6	111.8	112.1

[a]Conformer III (gas) undergoes transformation to the zwitterion form (sol) during DMol/COSMO geometry optimization in aqueous solvent.

The conformer I is stable in the solvent (see Table V) and the geometry changes are not significant. The solvent affects mainly the charge distribution of the molecule. The solvation energy of the structure I is 13.6 and 13.0 kcal/mol for VWN and BP Hamiltonians, respectively.

In the aqueous solvent the zwitterion is the most stable form of glycine and with respect to structure I, it is stabilized by about 10.7 and 8.7 kcal/mol with the VWN and B Hamiltonians, respectively. This is in good agreement with the experimental estimate of about 10 kcal/mol and previous AM1/COSMO[15] results.

The zwitterion structure exists only in the solvent environment. Using the DMol/COSMO optimized structure of zwitterion in the gas phase DMol energy calculations we find the zwitterion to be much less stable by 38.6 and 41.6 kcal/mol for VWN and BP calculations, respectively. This large electrostatic solvation energy is in qualitative agreement with AM1/COSMO results.[15]

The present study clearly indicates the need for accurate and efficient geometry optimization of the molecules in the presence of the solvent, as implemented in our DMol/COSMO approach. Without the gradient optimization capability we would not have been led to the zwitterion structure of the glycine, which is preferable in the aqueous environment.

V. CONCLUSION

We have developed the theory of DMol/COSMO energy and gradients. The formulation uses the projected density, as in the original DMol program, which improves the computational performance of the method. To our knowledge, this is the first implementation of the COSMO solvation model to calculate analytic gradients within the density functional method. The gradient formulation includes forces acting on both solute and cavity surface and allows for accurate optimization of the geometry of solute in the presence of solvent. As in the original COSMO model,[15] our implementation is more accurate in the case of solvents with high dielectric constants.

In the COSMO theory it is assumed that the solute charge density is completely contained within the cavity. In the DMol/COSMO approach we use electron density ρ defined on a numerical grid which extends beyond the cavity surface. Therefore a small amount of the solute charge appears outside the cavity. This charge defect problem does not exist in AM1/COSMO calculations[15] because atomic multipoles are used instead of the explicit electron density. The charge defect affects particularly the solvation energy of ions. This may partially account for too strong solvation of cations and too weak solvation of anions as observed in the paper. It appears that one can define a correction to the COSMO electrostatic potential which minimizes this charge defect.[48]

Much larger inaccuracies can arise from the choice of the cavity size or the estimation of the nonelectrostatic terms. The selection of the basis set and the DFT Hamiltonian is another source of possible errors. The systematic evaluation of these effects is currently under investigation in our lab and will be reported shortly.

Nevertheless, the present status of parametrization of the DMol/COSMO theory is already very satisfactory. The results reported in this paper are similar in accuracy to other implementations of the solvation models based on a dielectric continuum. Availability of the analytic gradients in DMol/COSMO program allows to study molecular conformations and chemical reactions in solution. Since the DMol/COSMO methodology leads to a very efficient computational implementation,[26] the present approach may be used to study transformations of large molecular systems in realistic solvent environment.

ACKNOWLEDGMENTS

[1] S. Miertus, E. Scrocco, and J. Tomasi, Chem. Phys. **55**, 117 (1981).
[2] J. Tomasi and M. Persico, Chem. Rev. **94**, 2027 (1994).
[3] E. L. Coitino, J. Tomasi, and O. Ventura, J. Chem. Soc. Faraday Trans. **90**, 1745 (1994).
[4] M. W. Wong, M. J. Frisch, and K. B. Wiberg, J. Am. Chem. Soc. **113**, 4776 (1991).
[5] 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).
[6] T. N. Truong and E. V. Stefanovich, Chem. Phys. Lett. (in press).
[7] A. Fortunelli and J. Tomasi, Chem. Phys. Lett. **231**, 34 (1994).
[8] J. L. Chen, L. Noodleman, D. A. Case, and D. Bashford, J. Phys. Chem. **98**, 11059 (1994).
[9] A. A. Rashin, M. A. Bukatin, J. Andzelm, and A. T. Hagler, Biophys. Chem. **51**, 375 (1994).
[10] R. J. Hall, M. M. Davidson, N. A. Burton, and I. H. Hillier, Phys. Chem. **99**, 921 (1995).
[11] K. Baldridge, R. Fine, and A. Hagler, J. Comput. Chem. **15**, 1217 (1994).
[12] C. J. Cramer and D. G. Truhlar, J. Am. Chem. Soc. **113**, 8305 (1991).
[13] C. J. Cramer and D. G. Truhlar, in *Reviews in Computational Chemistry,* edited by K. B. Lipkowitz and D. B. Boyd (VCH, New York, 1994).
[14] M. Szafran, M. M. Karelson, A. R. Katritzky, J. Kaput, and M. C. Zerner, J. Comput. Chem. **3**, 371 (1993).
[15] A. Klamt and G. Schuurmann, J. Chem. Soc. Perkin Trans. **2**, 799 (1993).
[16] D. Rinaldi, J.-L. Rivail, and N. Reguini, J. Comput. Chem. **13**, 675 (1992).
[17] R. Cammi and J. Tomasi, J. Chem. Phys. **101**, 3888 (1994).
[18] M. Basch, F. J. Luque, and M. Orozco, J. Comput. Chem. **15**, 446 (1994).
[19] H. Hoshi, M. Sakurai, Y. Inoue, and R. Chujo, J. Chem. Phys. **87**, 1107 (1987).
[20] E. L. Coitino, J. Tomasi, and R. Cammi, J. Comput. Chem. **16**, 20 (1995).
[21] B. Delley, J. Chem. Phys. **92**, 508 (1990); **94**, 7245 (1991).
[22] B. Delley, in *Modern Density Functional Theory: A Tool for Chemistry*, edited by P. Politzer and J. M. Seminario (Elsevier, New York, 1994).
[23] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, J. Chem. Phys. **71**, 3396 (1979).
[24] W. Kohn and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).
[25] R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules*, (Oxford, University, New York, 1989).
[26] To be published.
[27] DMol, version 950, Biosym Technologies, San Diego, CA 1995.
[28] S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
[29] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[30] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).
[31] J. Andzelm, J. Baker, A. Scheiner, and M. Wrinn, Int. J. Quant. Chem. (in press).
[32] J. Baker, J. Comput. Chem. **7**, 385 (1986); Optimize version 1.0 Biosym Technologies, San Diego, CA (1993).
[33] A. J. Bondi, Chem. Phys. **64**, 441 (1964).
[34] A. Ben-Naim and Y. Marcus, J. Chem. Phys. **81**, 2016 (1984).
[35] P. G. Jasien and G. Fitzgerald, J. Chem. Phys. **93**, 2554 (1990).

[36] *Handbook of Chemistry and Physics*, 71st ed. (CRC, Boca Raton, FL, 1990).

[37] R. G. Pearson, J. Am. Chem. Soc. **108**, 6109 (1986).

[38] W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab initio Molecular Orbital Theory* (Wiley, New York, 1986).

[39] C. Lim, D. Bashford, and M. Karplus, J. Phys. Chem. **95**, 5610 (1991).

[40] K. B. Wiberg and K. E. Laidig, J. Am. Chem. Soc. **109**, 5935 (1987).

[41] J. Gao and J. J. Pavelites, J. Am. Chem. Soc. **114**, 1912 (1992).

[42] J. Pranata, J. Comput. Chem. **14**, 685 (1993).

[43] V. Barone, C. Adamo, and F. Lelj, J. Chem. Phys. **102**, 364 (1995).

[44] R. Bonaccorsi, R. Cammi, and J. Tomasi, J. Am. Chem. Soc. **106**, 1945 (1984).

[45] H. S. Rzepa and M. M. Yi, J. Chem. Soc. Perkin Trans. **2**, 531 (1991).

[46] R. D. Suenram and F. J. Lovas, J. Am. Chem. Soc. **102**, 7180 (1980).

[47] F. Sim, A. St-Amant, I. Papai, and D. R. Salahub, J. Am. Chem. Soc. **114**, 4391 (1992).

[48] A. Klamt and V. Jonas (to be published).