Dimerization of Formamide in Gas Phase and Solution. An Ab Initio MC−MST Study

Carles Colominas,[†,‡] F. Javier Luque,[*,§] and Modesto Orozco[*,†]

Departament de Bioquímica, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain, Departament de Química Orgànica, Institut Químic de Sarria, Universitat Ramon Llull, Via Augusta 309, Barcelona 08017, Spain, and Departament de Fisicoquímica, Facultat Farmàcia, Universitat de Barcelona, Avgda Diagonal sn, Barcelona 08028, Spain

Received: March 8, 1999; In Final Form: May 10, 1999

Dimerization of formamide in the gas phase has been studied by a combination of high level quantum mechanical calculations (ab initio and density functional calculations) and Monte Carlo simulations. The influence of the solvent on the dimerization has been introduced by means of self-consistent reaction field calculations (using the Miertus−Scrocco−Tomasi formalism), as well as by the newly developed Monte Carlo−MST methodology. A complete description of the configurational map of the formamide dimer in aqueous and chloroform solution is provided. The large effect of solvation in the dimerization is clearly shown.

Introduction

Amide dimerization is a chemical process relevant to understanding the formation of protein structures and many processes of molecular recognition both in polar and apolar solvents.[1] In fact, the dimerization mode of cis amides is probably similar to the H-bonding pattern responsible for recognition of nucleic acid bases. Owing to the chemical and biochemical implications of amide−amide recognition, a large research effort has been conducted to determine the free energy of dimerization of cis and trans amides in different solvents.[2] In most cases the association was examined from the changes in the infrared spectra of amide monomers and dimers.[2] There is also a large number of theoretical studies on this topic,[3] which were performed to determine accurately the free energy of dimerization of amides in the gas phase, and also to estimate the effect of the environment in the dimerization.

Most models of amide dimerization assume that this interaction occurs mainly by H-bonding, especially for the cis amides. In fact, experimental measures detect the dimerization mainly as a change in N−H stretching frequencies, which are expected to be greatly altered only upon H-bonding. Nevertheless, recent theoretical calculations have questioned that assumption, since the results indicate that amide dimerization in water, if any, is not modulated by formation of H-bonds, but by other less specific interactions such as partial stacking.[3b]

In this paper we reinvestigate the issue of amide dimerization in the gas phase, apolar and polar solvents using a combination of state-of-the-art methods for quantum mechanical calculations and classical simulations. To achieve the highest levels of theory in calculations, the smallest amide−amide system, i.e., the formamide dimer, has been considered. The results are useful to understanding *cis*-amide dimerization. Furthermore, since accurate experimental measures[3b,c,l] have demonstrated that the association constant of cis and trans amides is very similar for solvents with polarity equal or superior to that of chloroform, the results can be also valuable to understanding better *trans*-amide dimerization in polar solvents.

Methods

Quantum Mechanical Gas-Phase Calculations. Preliminary Monte Carlo (MC) calculations combined with Powell and Simplex optimizations were used to explore the configurational space of the formamide dimer in gas phase. Five energy minima (structures 2a−d and 2f in Figure 1) were located. Subsequently, these minima were optimized and characterized as true energy minima from B3LYP/6-31G(d) frequency analysis. Previous calculations in structurally related systems[4] and comparison with G2 results (see below) confirm the reliability of the B3LYP/6-31G(d) estimates. Another structure (2e in Figure 1), which corresponds to a single H-bond dimer and which seems to be quite populated in MC runs, was also considered at the QM level. The corresponding stationary point obtained after geometry optimization was found, nevertheless, to be not a real energy minimum in the frequency analysis.

The most stable configuration of the dimer in gas phase as determined from B3LYP/6-31G(d) calculations was then studied using a variety of theoretical methods ranging from HF/6-31G(d) to G1, G2 and G2MP2[5] calculations, including geometry optimization at the respective levels (see Table 1). In all cases the dimerization energy was corrected for the basis set superposition error (BSSE) using the counterpoise method.[6] Dimerization enthalpies and entropies were determined using the standard procedures in *Gaussian 94*[7] and considering a reference state corresponding to an ideal gas at 1 atm and 298 K.

Force-Field Parametrization. A set of parameters was derived to ensure a correct energetics for the amide−amide dimerization. The van der Waals parameters were taken from the OPLS force field,[8] while atomic charges were determined by fitting to electrostatic potentials[9] computed at the B3LYP/6-31G(d) level.[10] At this point, it must be noted that ESP charges determined at this level of theory are close to those derived at the CI level.[11] The final charges were then slightly adjusted

* Correspondence to M. Orozco or F. J. Luque.
† Departament de Bioquímica, Universitat de Barcelona.
‡ Universitat Ramon Llull.
§ Departament de Fisicoquímica, Universitat de Barcelona.

10.1021/jp9908092 CCC: $18.00
Published on Web 07/16/1999

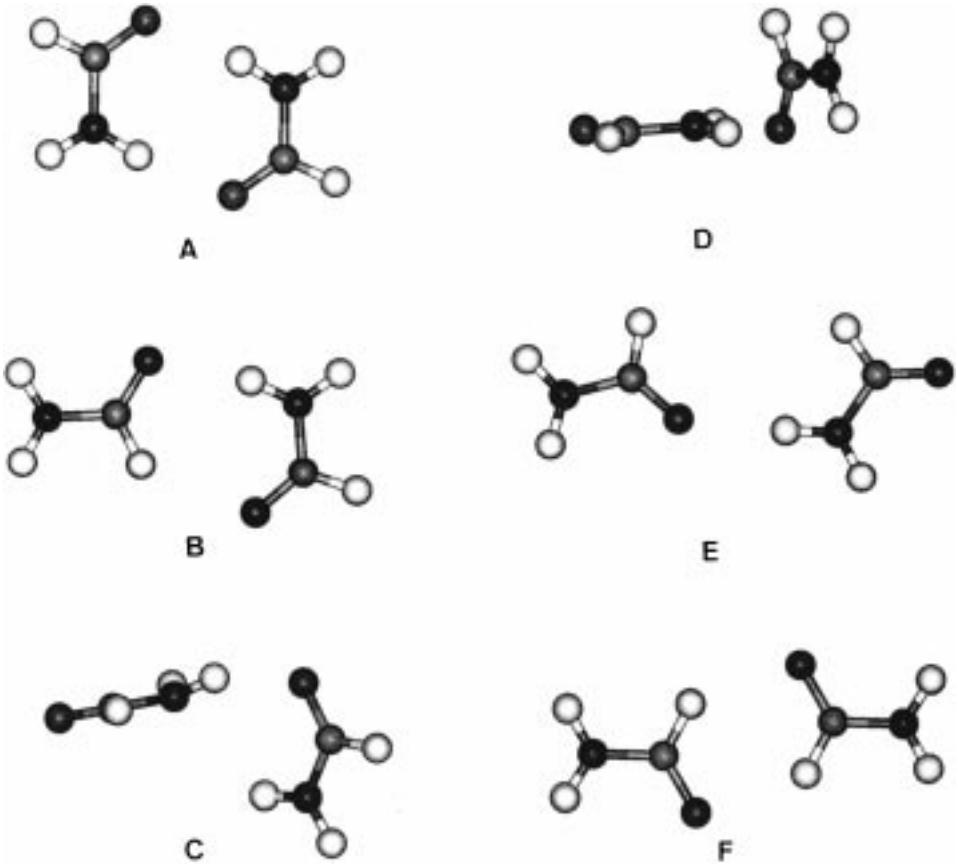

Figure 1. Structures of the six formamide dimers considered in DFT gas-phase calculations.

TABLE 1: Interaction Energy (BSSE-corrected), Enthalpy, and Free Energy for the Double H-Bond Dimer of Formamide Determined at Different Levels of Theory in the Gas Phase[a]

geometry	method	ΔE	ΔH	ΔG
HF/6-31G(d)	HF/6-31G(d)	−11.11	−9.10	2.05
MP2/6-31+G(d)	HF/6-31+G(d)	−10.95	−8.94	2.22
MP2/6-31+G(d)	MP2/6-31+G(d)	−12.52	−10.50	0.65
MP2/6-31+G(d)	HF/6-31++G(3df,3pd)	−10.54	−8.53	2.62
MP2/6-31+G(d)	MP2/6-311++(3df,3pd)	−13.65	−11.64	−0.49
MP2/6-31+G(d)	MP2/aug-cc-pVTZ	−14.50	−12.49	−1.34
MP2(full)/6-31G(d)	MP2(full)/6-31G(d)	−12.42	−10.41	0.75
MP2(full)/6-31G(d)	MP2/6-311G(d,p)	−10.73	−8.71	2.44
MP2(full)/6-31G(d)	MP4/6-311G(d,p)	−10.49	−8.48	2.68
MP2(full)/6-31G(d)	MP2/6-311+G(d,p)	−11.12	−9.11	2.04
MP2(full)/6-31G(d)	MP4/6-311+G(d,p)	−10.95	−8.94	2.21
MP2(full)/6-31G(d)	MP2/6-311G(2df,p)	−12.55	−10.53	0.62
MP2(full)/6-31G(d)	MP4/6-311G(2df,p)	−12.30	−10.29	0.87
MP2(full)/6-31G(d)	QCISD(T)/6-311G(d,p)	−10.72	−8.71	2.44
MP2(full)/6-31G(d)	MP2/6-311+G(3df,2p)	−13.58	−11.57	−0.42
B3LYP/6-31G(d)	B3LYP/6-31G(d)	−13.35	−11.23	1.04
B3LYP/6-31G(d)	B3LYP/6-31G(d)	−13.35	−11.34	−0.19
MP2(full)/6-31G(d)	G1		−10.99	0.16
MP2(full)/6-31G(d)	G2(MP2)		−11.63	−0.48
MP2(full)/6-31G(d)	G2		−11.57	−0.42

[a] A reference state of ideal gas at 298 K and 1 atm was used for thermodynamic calculations. HF/6-31G(d) frequencies were scaled following the G2 procedure. MP2(full) means that no frozen-core approximation was used. All of the values are in kcal/mol. [b]B3LYP/6-31G(d) frequencies.

to improve the fitting between classical and B3LYP/6−31(d) interaction energies.

Following the Monte Carlo Miertus−Scrocco−Tomasi (MC−MST) procedure,[12] a dual set of charges fitted to both the electrostatic potential and field (ESPF charges)[13] is also required to treat the solute−solvent interactions including the mutual polarization terms. These charges were determined using our standard ESPF strategy[13] at the HF/6-31G(d) level and considering both water and chloroform as solvents.

Solvation Calculations. Solvent effects were introduced at the QM level following a HF/6-31G(d) optimized version[14] of the well-known Miertus, Scrocco, and Tomasi algorithm (MST;[12] see eq 1). In the MST framework the electrostatic contribution (ΔG_{ele}) was determined using the PCM methodology (eq 2), cavitation (ΔG_{cav}) was computed following the Pierotti−Claverie[15] formalism, and the van der Waals term (ΔG_{vW}) was evaluated using an atom-optimized linear relationship with solvent-accessible surface.[14] The vacuum optimized geometries were used in all cases.

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{vW} + \Delta G_{ele} \quad (1)$$

$$\Delta G_{ele} = \left\langle \Psi^{sol} \left| \hat{H}^0 + \frac{1}{2}\hat{V}_R^{sol} \right| \Psi^{sol} \right\rangle - \langle \Psi^0 \mid \hat{H}^0 \mid \Psi^0 \rangle \quad (2)$$

where V_R refers to the perturbational operator representing the solvent reaction field, and indexes 0 and sol stand for gas phase and solution states.

MC−MST calculations were used to explore the configurational space of the formamide dimer in gas phase, chloroform, and aqueous solution. Such an exploration permits to examine the nature of the dimerization and to derive thermodynamic data for such process. MC−MST calculations rely on the quasiclassical formalism of the MST algorithm,[13] where the electrostatic term is determined from eq 3, which can be rigorously derived from perturbational theory.[16] Equation 3 allows for representing the electrostatic contribution to solvation (including polarization effects) using a Coulombic expression, which speeds up the

calculation and makes it possible to enlarge the simulation while keeping a reasonable computational cost.[17]

$$\Delta G_{ele} = \frac{1}{2}\sum_{i=1}^{N} Q_i^0 V_\sigma^{sol} \tag{3}$$

In eq 3, Q_i^0 stands for the set of point charges that represent the gas-phase charge distribution and V_σ^{sol} is the electrostatic potential generated by the solvent apparent charges spread over the solute/solvent interface (they are generated in response to the presence of the charge distribution of the fully polarized solute in solution; see refs 14,16 for more details). Note that eq 3 requires knowledge of two sets of charges for the solute: one (Q_i^0) represents the gas-phase charge distribution in eq 3, whereas the other (Q_i^{sol}) describes the polarized charge distribution in solution, which generates the solvent reaction field denoted by V_σ^{sol}. These sets of charges are determined following the ESPF strategy (see above) as explained in detail elsewhere.[14]

Combination of eq 3 with a standard force field allows for defining a mean energy functional as that shown in eq 4. Note here that another set of effective charges (Q_i) is required to reproduce properly solute−solute interactions (see above).

$$E = \Delta G_{sol} + \sum_{i,i'} \frac{Q_i Q_{i'}}{R_{ii'}} + \sum_{i,i'} \left[\frac{A_{ii'}}{R_{ii'}^{12}} - \frac{C_{ii'}}{R_{ii'}^{6}} \right] \tag{4}$$

where indexes i and i' refer to atoms in the two monomers and A, C are standard van der Waals parameters (see above).

The Metropolis algorithm is used to generate a Boltzmann ensemble of configurations of the dimer at a certain concentration.[17] Accordingly, configurations of the system are randomly generated by translating and rotating one monomer with respect to the other. The mean energy (eq 4) is then computed and the configuration is accepted or rejected following the Metropolis rules. MC−MST simulations were typically performed using a multiple copy approach, which means that several copies of one monomer (A) are placed randomly around a central one (B). Each monomer A interacts with the central monomer B but is unable to see other copies or the solvent reaction field generated by them. This approach gives results less dependent on the starting configuration than those determined from single copy MC runs.[17] In this study we used 40 copies for simulations in water and chloroform and 20 copies for simulations in gas phase. Each copy generates an independent Markow chain that was followed for 30 000 (water and chloroform) or 160 000 (gas phase) configurations. This implies MC runs of 1.2 M configurations for simulations in water and chloroform and 3.2 M configurations for simulations in gas phase.

The accepted configurations were classified in different geometrical families to facilitate the analysis. Configurations were divided in two groups depending on whether the distance between the centers-of-mass of the interacting monomers is greater than a cutoff radius (r_{cut}), which was defined (see below) as the value where the radial distribution function is 1.0 for gas phase (r_{cut} = 5.1 Å) and chloroform (r_{cut} = 5.4 Å), and where it is a minimum in aqueous solution (r_{cut} = 5.6 Å). Configurations within the cutoff radii were grouped in three categories: overlapped, H-bonded, and nonoverlapped. The overlapped category meets those configurations in which any atom of monomer A lies above/below the rectangle defined by monomer B in the molecular plane. The overlapped category is divided into three families according to the angle (α) between the two formamide planes: stacking ($\alpha < 35$ degrees), T-shape ($\alpha > 65$ degrees), and other ($35 < \alpha < 65$ degrees). The H-bonded category contains those configurations where the distance between oxygen and nitrogen atoms is less than 3.5 Å and the O−H−N angle is greater than 120 degrees. This category is partitioned into single and double H-bonded configurations. Finally, the nonoverlapped category contains all configurations within r_{cut} not included in the two preceding categories.

Inspection of the configurational space sampled after equilibration allowed us to estimate the dimerization free energy, which at a given monomer can be determined from eq 5. In this equation V^0 refers to the volume of the box necessary to have 1 M concentration for each of the two monomers (2 M in formamide in our study) and V^{box} is the volume of the box used in the simulation. In this study the box size was chosen as to have 1 M concentration of formamide, which means that the second term in eq 5 amounts to −0.4 kcal/mol.

$$\Delta G_{dim} = -RT \ln\left(\frac{N_{dimer}}{N_{monomer}}\right) + RT \ln\left(\frac{V^0}{V^{box}}\right) \tag{5}$$

where N_{dimer} and $N_{monomer}$ are the number of Metropolis accepted configurations which correspond to dimer (N_{dimer}) and separated monomers ($N_{monomer}$).

Computational Details. Ab initio and density functional calculations were performed with *Gaussian 94*.[7] ESP and ESPF charges were determined using MOPETE/MOPFIT programs.[18] Ab initio SCRF calculations were carried out using a locally modified version of *MonsterGauss*.[19] Monte Carlo−MST calculations were performed using our MC−MST code.[20] Calculations were performed in the Origin-2000 and SP2 computers of the Centre de Supercomputació de Catalunya (CESCA) and in workstations in our laboratory.

Results and Discussion

Quantum Mechanical Gas-Phase Calculations. The symmetrical double H-bond dimer is expected to be the major form in gas phase. Thus, our first aim was to examine this complex at different levels of theory to obtain accurate thermodynamic quantities of the dimerization and to assess the accuracy of lower level calculations. Table 1 gives the dimerization energies, enthalpies, and free energies for the double H-bonded dimer. Calculations ranged from HF or B3LYP/6-31G(d) methods to state-of-the-art G2 or MP2/aug-cc-pVTZ calculations. The geometries were optimized at four levels: HF/6-31G(d), MP2/6-31+G(d), MP2(full)/6-31G(d), and B3LYP/6-31G(d). The reference state was the ideal gas at 298 K and 1 atm. Following the G2 formalism, entropic and thermal corrections were determined using scaled HF/6-31G(d) frequencies. However, unscaled B3LYP/6-31G(d) frequencies were also considered in B3LYP calculations.

Inspection of the results in Table 1 shows that dimerization is largely favored by enthalpic factors and disfavored by entropic terms. The final result is that the free energy of dimerization is not far from zero. Thus, the dimerization free energy is −0.5 kcal/mol at the G2 level, while it amounts to −1.3 kcal/mol from MP2/aug-cc-pVTZ calculations. Lower level calculations provide often positive values, which does not seem to agree with chemical intuition.

As expected,[4] extension of the basis set is crucial for a correct description of the interaction energy at the HF, MPx, or CI levels

(our previous results[4] with related systems indicate that this does not occur for DFT calculations). Inclusion of polarization functions is particularly relevant, whereas diffuse functions have little effect.[4] Thus, extension from 6 to 311G(d,p) to 6-311G-(2df,p) to 6-311+G(3df,2p) changes the binding energy by almost 2 and 1 kcal/mol. Further extension to 6-311++G-(3df,3pd) does not lead to relevant changes. Electron correlation effects are necessary to describe properly the dimerization. Thus, the binding energy increases by almost 3 kcal/mol from HF to MP2 levels when the 6-311++G(3df,3pd) basis is used. Most of the electron correlation effect is accounted for at the MP2 level, which suggests that more elaborated electron correlation methods do not seem advisable. Finally, the results point out the reliability of the B3LYP method to estimate the dimerization energy in H-bonded systems. It is worth noting that the accuracy of present DFT results is sensibly better than those obtained with more expensive HF or MPx calculations.

Our theoretical estimates can be compared with previous high-level theoretical studies. Thus, Suhai[3g] reported a dimerization energy of −7 kcal/mol at the MP4 level using a DZ(d,p) basis, which underestimates the strength of the interaction. Neuheuser et al.[3d] reported a binding energy of −11.4 kcal/mol at the MP2 level using a DZP basis, which underestimates by at least 2 kcal/mol the stability of the dimer according to higher level calculations. Dixon et al.[3e] obtained an enthalpy difference of −12.3 kcal/mol after full BSSE correction at the MP2/aug-cc-pVDZ level, a value in agreement with our best estimates. Florian and Johnson[3f] reported MP2/6-31G(d,p) and different DFT estimates of the dimerization enthalpy. While MP2 and S-VWN values are clearly incorrect, the B-LYP functional, especially when the 6-31G(d) basis is used, provides more reasonable values. Finally, Hobza and Sponer[3h] reported dimerization energies around −12.4 kcal/mol at the MP2 and CCSD-(T) levels using the aug-cc-pVTZ/s(cc-pVDZ) basis, confirming the small gain in accuracy arising from an increase in the level of theory beyond the MP2 level.

Comparison with experimental results is more difficult, especially owing to the lack of accurate measures of gas-phase amide dimerization. A rough estimate can be derived from the dimerization of cis amides in very apolar solvents. There is a notable dispersion in the results, but in general dimerization free energies of around −3 kcal/mol have been reported in CCl_4 (see ref 2l and references therein), even though the value depends greatly on the *cis*-amide system.[2d,f,j,m] Data in other very apolar solvents are more scarce, but IR experiments[2f] suggest that the dimerization is around 0.6 kcal/mol stronger in cyclohexane than in CCl_4, and that benzene hinders the dimerization by around 1 kcal/mol with regard to CCl_4. In summary, experimental data on related systems indicate that the gas-phase dimerization free energy ranges between −3 to −4 kcal/mol. Let us stress again that this is a rough estimate derived from dimerization data in very apolar solvents of large cis amides, which can likely establish interactions other than double H-bonding. Comparison of the theoretical estimates should be then made with caution. Since the "experimental" value is obtained for a 1 M reference state, while theoretical results refer to an ideal gas at 298 K and 1 atm, correcting the theoretical values for the change in reference state yields estimates of −2.4 (G2), −2.1 (B3LYP/6-31G(d)), and −3.2 (MP2/aug-cc-pVTZ) kcal/mol for a 1 M reference state. The agreement with experimental measures is then remarkable considering the large range of uncertainty for both theoretical and "experimental" estimates.

TABLE 2: Dimerization Energy (BSSE-corrected), Enthalpy and Free Energy for Different Conformations of the Formamide Dimer in the Gas Phase[a]

dimer	ΔE	ΔH	ΔG
a	−13.4	−11.2	1.0
b	−8.3	−6.5	4.6
c	−6.7	−4.8	6.0
d	−6.0	−4.2	4.9
e	−6.7		
f	−3.2	−1.6	7.5

[a] B3LYP/6-31G(d) geometries, energies, and frequencies were used. A reference state of ideal gas at 298 K and 1 atm was used for thermodynamic calculations. All of the values are in kcal/mol. See Figure 1 for the structure of the dimers.

TABLE 3: Interaction Energies (kcal/mol) Computed at the B3LYP/6-31G(d) and Classical Levels for B3LYP and Classically Optimized Geometries[a]

dimer	geometry	ΔE (B3LYP)	ΔE (classical)
a	B3LYP	−13.4	−13.5
a	classical		−13.0
b	B3LYP	−8.3	−8.6
b	classical		−8.4
c	B3LYP	−6.7	−8.1
c	classical		−7.8
d	B3LYP	−6.0	−6.2
d	classical		−7.7
e	B3LYP[b]	−6.7	−6.2
f	B3LYP	−3.2	−4.6
f	classical		−4.6

[a] The amide planarity was fixed in classical intermolecular geometry optimizations. [b]Not a real minimum in the B3LYP/6-31G(d) potential energy hypersurface; the structure is not detected as a stationary point in classical optimizations.

Other dimerization modes in gas phase were explored at the B3LYP/6-31G(d) level owing to the good performance and relative inexpensiveness of this method. For this purpose, in a preliminary step the gas-phase configurational space was explored with Monte Carlo and minimization techniques, leading to five energy minima (structures a−d and f in Figure 1), which were further refined and characterized by frequency analysis at the B3LYP/6-31G(d) level. The analysis also included structure e in Figure 1, which is not a real minimum but was found to be representative of a family of structures widely sampled in gas-phase MC simulations of the dimer. Interaction energies (BSSE-corrected), enthalpies, and free energies of dimerization computed from B3LYP/6-31G(d) unscaled frequencies are given in Table 2. The results indicate that the double H-bond is more stable that any other minimum, in agreement with previous calculations.[3d] Interestingly, the dimerization energy of structure e is similar or even better than that of other structures, which agrees with its large population in MC calculations (see below). This suggests that assuming that the gas-phase configurational space of formamide can be described considering only the energy minima can lead to erroneous results, since regions stable energetically can be very populated and contribute to the dimerization even though they are not true minima in the potential energy hypersurface.

Monte Carlo simulations are expected to be largely dependent on force-field parameters. B3LYP/6-31G(d) dimerization energies are then valuable to check the accuracy of the solute−solute force field. Table 3 gives the corresponding energies for the six structures in Figure 1 optimized at the B3LYP/6-31G-(d) and classical levels. The classical force field developed here performs well, not only regarding the double H-bonded dimer but also the rest of structures. For the most stable minima the

TABLE 4: MST-HF/6-31G(d) Estimates of the Free Energy of Solvation (kcal/mol) and Its Components in Water and Chloroform[a]

structure	G^{ster}	ΔG^{ele}	ΔG_{solv}	$\Delta\Delta G_{solv}$
		water		
formamide	2.3	−11.7	−9.4	
dimer a	4.2	−13.8	−9.5	9.3
dimer b	4.2	−16.8	−12.6	6.2
dimer c	4.1	−17.7	−13.6	5.2
dimer d	4.1	−20.0	−15.9	2.9
dimer e	5.1	−20.1	−15.9	2.9
dimer f	3.9	−19.2	−15.3	3.5
		chloroform		
formamide	2.8	−4.0	−6.7	
dimer a	4.5	−4.4	−8.9	4.5
dimer b	4.6	−5.2	−9.7	3.7
dimer c	4.7	−5.7	−10.4	3.0
dimer d	4.6	−6.9	−11.5	1.9
dimer e	4.8	−7.2	−11.9	1.5
dimer f	4.6	−5.9	−10.5	2.9

[a] Relative values ($\Delta\Delta G_{solv}$) are determined from the difference between the free energy of hydration of the dimer and the two monomers with the same intramolecular geometry.

TABLE 5: Free Energies of Solvation in Water and Chloroform of the Six Dimers of Formamide Shown in Figure 1[a]

solvent	dimer	ΔG_{solv}(QM)	ΔG_{solv}(sc)[b]	ΔG_{solv}(sc)[c]
water				
	a	−9.5	−9.2	−9.0
	b	−12.6	−12.7	−12.4
	c	−13.6	−13.0	−13.1
	d	−15.9	−14.8	−13.2
	e	−15.9	−14.6	
	f	−15.3	−16.2	−16.2
chloroform				
	a	−8.9	−8.6	−8.3
	b	−9.7	−9.6	−9.5
	c	−10.4	−10.0	−10.4
	d	−11.5	−10.7	−10.6
	e	−11.9	−11.0	
	f	−10.5	−10.8	−10.8

[a] QM stands for the ab initio 6-31G(d)-MST estimates and sc for the equivalent semiclassical values. [b]Dimer fully optimized at the B3LYP level. [c]Dimer optimized at the classical level keeping the planarity of formamide.

force field reproduces within 0.1 kcal/mol the B3LYP estimates, which in turn differ as much as 0.3 kcal/mol from G2 values. Considering the structures, the RMS error is 0.8 kcal/mol and the largest deviation is 1.4 kcal/mol. Indeed, similar interaction energies are found when the intermolecular geometry is optimized classically (RMS deviation: 1.1 kcal/mol; maximum error: 1.7 kcal/mol), even though the planarity of the amide was fixed in these calculations. As a comparison, when OPLS parameters in BOSS3.4[8] are used, the RMS deviation from the B3LYP/6-31G(d) values is 1.6 kcal/mol and the largest error is 3.4 kcal/mol.

QM−SCRF Representation of Solvation. Our HF/6-31G-(d) optimized version of the MST method was used to examine the solvent effect in the dimerization of formamide. For this purpose the free energies of solvation (ΔG_{solv}) in water and chloroform of formamide and of the six dimer structures shown in Figure 1 were determined. The results are shown in Tables 4 (water) and 5 (chloroform).

Both water and chloroform hinder the formation of the dimers found as the most stable ones in gas phase, as noted in the positive values of $\Delta\Delta G_{solv}$. As expected, the electrostatic term disfavors dimerization, while the steric term favors the process in the two solvents. The disturbing effect of water is larger than that of chloroform, but it is quite large even in this latter solvent, with $\Delta\Delta G_{solv}$ values ranging from 2 to 5 kcal/mol. Since the most stable gas phase dimers are generally those leading to large annihilation of the monomer multipole moments, the dimers are very apolar and are greatly disfavored upon solvation. Thus, the double H-bond dimer is destabilized by around 10 and 5 kcal/mol in water and chloroform, whereas dimer e is disfavored only by around 3 and 2 kcal/mol, respectively. Overall, the formation of any of the dimers shown in Figure 1 is disfavored in the two solvents, as stated from comparison of the $\Delta\Delta G_{solv}$ values (Table 4) with the gas phase dimerization free energies (Table 2; note the change in reference state). Therefore, the dimerization, if any, must involve structures other than those found in gas phase, and accordingly little insight about the dimerization is gained from the analysis of the most stable gas-phase dimers.

MST-HF/6-31G(d) calculations of amides are expected to be quite accurate (for instance, the ΔG_{solv} of acetamide is reproduced with an error of 0.4 and less than 0.1 kcal/mol in water and chloroform). This suggests that ab initio data can be used to check the semiclassical free energies of solvation computed in MC−MST calculations. These values are −10.0 and −7.0 kcal/mol for the planar formamide in water and chloroform (−10.4 and −7.1 kcal/mol when the B3LYP-optimized geometry is considered). These values compare well with the MST-HF/6-31G(d) results, which are respectively −9.9 and −6.9 kcal/mol (−9.4 and −6.7 kcal/mol for the B3LYP optimized geometry).

Table 5 compares QM and semiclassical free energies of solvation in water and chloroform of the formamide dimers shown in Figure 1. The agreement between QM and semiclassical results is excellent when the same geometry is considered (RMS errors of 0.8 and 0.5 kcal/mol in water and chloroform). Small discrepancies in ΔG_{solv} occur when classical geometries with the planar formamide are utilized (RMS errors of 0.7 and 0.2 kcal/mol in water and chloroform). In summary, despite the simplicity of the semiclassical method, this approach gives reliable free energies of solvation for the monomer and dimers. In conjuction with data in Table 3, these results support the quality of the classical (solute−solute)−semiclassical (solvation) approach used in MC−MST calculations.

MC−MST Calculations. The dimerization of formamide in gas phase, chloroform, and water was explored by using MC−MST calculations. As seen in Figure 2, the simulations in chloroform and water seem well equilibrated after few hundreds of configurations for each copy, while the gas-phase simulation needs a longer equilibration period (around 30K for each copy) because of the stiffness of the gas-phase energy hypersurface. After equilibration, the accepted configurations were grouped into different categories to analyze the configurational space and to estimate dimerization free energies, which can then compared with available experimental data in related systems.

Gas Phase. The results clearly indicate that the bound dimer is the major form (see Table 6). Thus, around 97% of the accepted configurations correspond to structures where the distance between the centers-of-mass of the monomers are lower than 5.1 Å. Most of them are H-bonded structures, which account for around 94% of the accepted configurations. From these values, if one treats as "bound" dimers those structures having a separation between centers-of-mass of monomers smaller than r_{cut}, a dimerization free energy around −2.5 kcal/mol is obtained. Alternatively, if one considers as "bound"

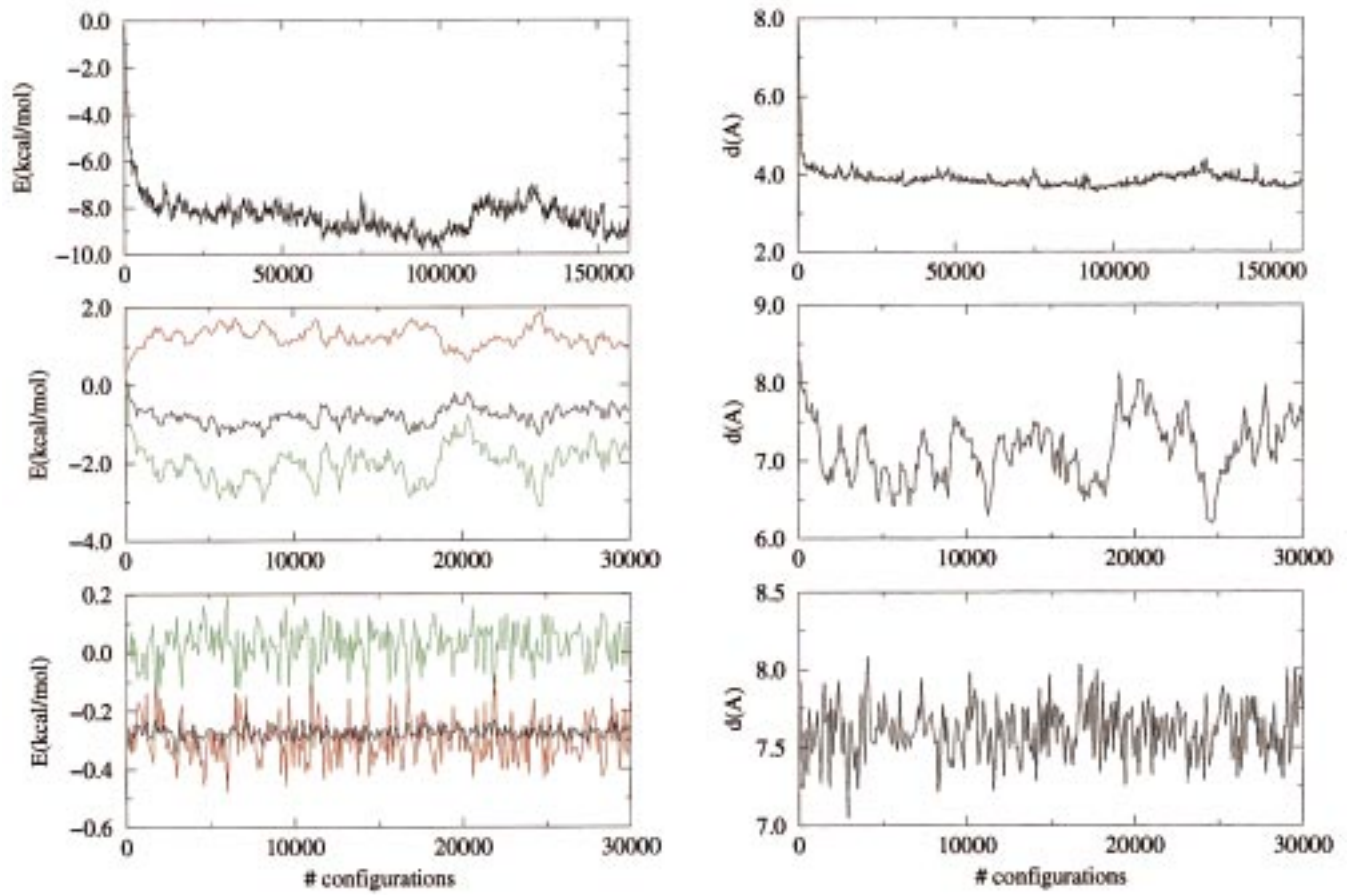

Figure 2. Left: Energy (total, black; solute−solute, green; and solvation, red) changes determined in gas phase (top), chloroform (middle), and water (bottom). Right: Distance between the centers-of-mass of the two formamides in gas phase (top), chloroform (middle), and water (bottom). In all cases the values correspond to the average of all of the copies in multiple-copy MC−MST calculations.

TABLE 6: Occurrence (in percentage) of Different Conformations of the Formamide Dimer During the MC−MST Calculations[a]

solvent	$r < r_{cut}$	stack	T-shape	mixed	other	1 H-bond	2 H-bonds
gas phase	97.20	0.06	0.40	0.30	2.81	34.10	59.53
chloroform	36.41	0.18	0.78	0.84	7.82	23.96	2.83
water	13.77	0.35	1.72	1.51	9.38	0.81	0.00

[a] Here, r_{cut} is defined from radial distribution functions (see text) and is 5.1 (gas), 5.4 (chloroform), and 5.6 Å (water). See text for definition of the different "families".

dimers those H-bonded configurations, a definition likely closer to what is detected in IR experiments, the free energy of dimerization is around −2.0 kcal/mol. In any case, the agreement with the "experimental" estimate (see above) derived from dimerization data in very apolar solvents is remarkable. Likewise, the free energy change for the approach of two monomers from infinite to form a double H-bond dimer is estimated to be −2.2 kcal/mol. This value compares very well with QM data (−2.4 (G2); −2.1 (B3LYP); −3.2 (MP2/augcc-pVTZ) kcal/mol) examined before.

An interesting point in MC−MST calculations is the large population of single H-bonded structures, which represent more than a third of the total H-bonded dimers. In fact, the last snapshot of the MC simulation clearly reflects this feature (see Figure 3A). This is also noted in the radial distribution function (Figure 4A) where single H-bonded structures are responsible for the second peak, while the first one is dominated by double H-bonded forms. The large population of single H-bonded structures is surprising in view of QM results, which found that the double H-bonded dimer was more stable than any single H-bonded minimum. This finding has a clear entropic origin and can be rationalized from the density plots in Figure 5. Thus, the double H-bonded dimer occupies a very narrow region of the configurational space. However, there are many possible single H-bond structures (bound to each (N)H or (C)O) having smaller energy, even though they are not true energy minima. A a result, the population of single H-bonded dimers is larger than expected from the differences in stability. Therefore, caution is necessary in the interpretation of QM free energies derived for a limited number of minima.

Chloroform. Simulations in chloroform provide a picture of the formamide dimerization different from the gas-phase situation, demonstrating that chloroform cannot be considered to be a very apolar solvent. Only around 36% of the accepted configurations correspond to structures having a distance between the centers-of-mass less than 5.4 Å, which would lead to a dimerization free energy of c.a. −0.1 kcal/mol. If only H-bonded structures are considered as "bound" dimers, the free energy of dimerization is 0.2 kcal/mol. In summary, MC−MST calculations indicate that the dimerization in chloroform leads to negligible changes in free energy. In fact, these values are in agreement with experimentally measured dimerization constants ($K = 1-3$ M^{-1}) for cis amides in chloroform.[2l]

The nature of the "bonded" forms also changes dramatically with respect to the gas-phase situation (see Figures 3B-5B and Table 6). Whereas the double H-bonded dimer is the dominant structure in gas phase and the population of other structures

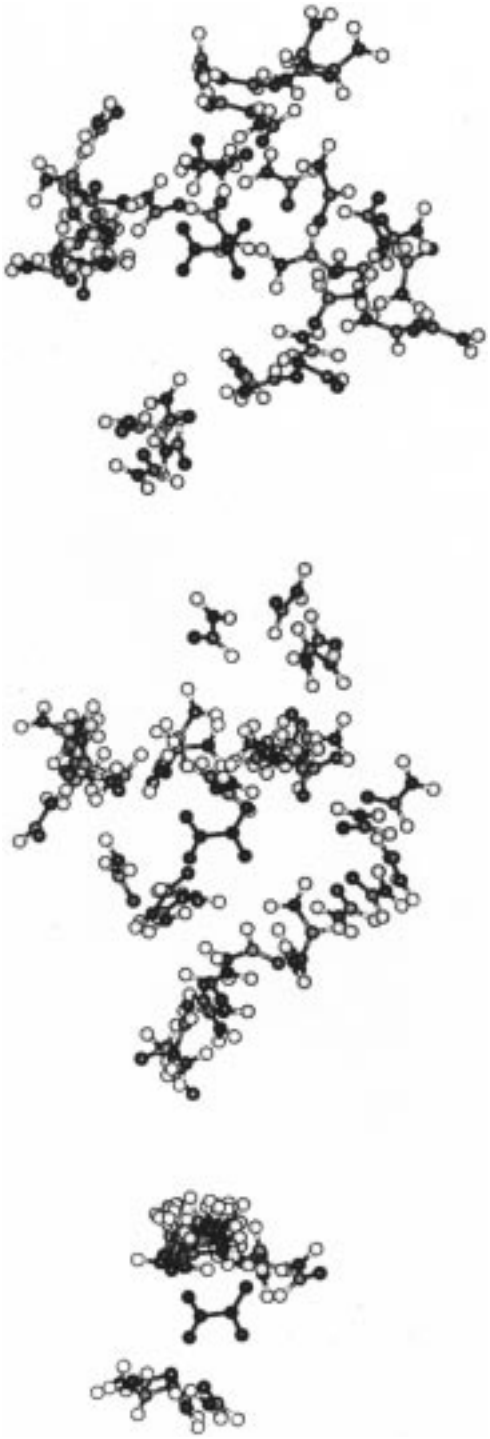

Figure 3. Representation of the last snapshot of MC−MST simulations in gas phase (bottom), chloroform (middle), and aqueous solution (top).

was very small, in chloroform the double H-bonded form is minor (only 2.83% of the accepted configurations) and the single H-bonded structures correspond to nearly 24%, they being the dominant "bonded" dimer. Note also that the importance of non-H-bonded dimers increases, especially those corresponding to non-well-defined structures. These findings agree qualitatively to those obtained by Jorgensen[3b] using discrete MC−PMF calculations of the dimerization of NMA in chloroform.

The change in the relative stability of single and double H-structures, which is clear in the radial distribution function (Figure 4B) and the density contour (Figure 5B), seems surprising. However, it can be realized considering QM−SCRF results (Table 4), which show that the cyclic dimer is disfavored

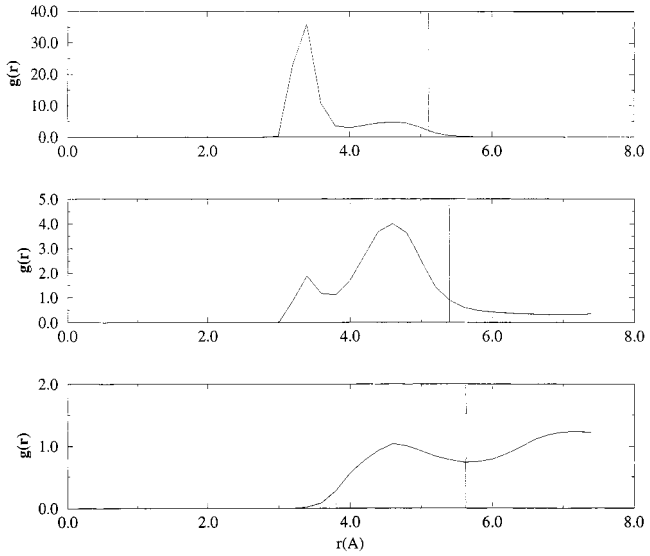

Figure 4. Formamide−formamide (center-of-mass−center-of-mass) radial distribution functions for MC−MST calculations in gas phase (top), chloroform (middle), and water (bottom). The value derived as r_{cut} is represented as a straight line in each case.

by chloroform in 3−4 kcal/mol with regard single H-bonded structures ($\Delta\Delta G_{solv}$ in Table 4). The fact that single H-bonded structures are around 10 times more populated than double H-bonded dimers suggests that no large differences in the dimerization free energies of cis and trans amides are expected in chloroform. In fact, the experimental values of cis and trans amides are identical within the experimental error,[2l] while there are large differences in solvents such as CCl_4. This supports the results derived from MC−MST calculations.

Water. As expected from QM−SCRF calculations (Table 4), the configurational space of the formamide dimer in water is different from those in gas phase and chloroform. Thus, only 13.8% of the accepted configurations have distances between centers-of-mass lower 5.6 Å, and most of these configurations correspond to non-well-defined structures (Table 6), T-shape, and stacking conformations, while single H-bonded forms occur in only 0.8% of the accepted configurations, and the double H-bond dimer is not detected. A complete view of the configurational space is given by Figures 3C−5C. Clearly, bonded configurations are rare, and when the two formamide molecules are close, they do not adopt well-defined conformations. The radial distribution function (Figure 4C) shows the disappearance of the first peak, corresponding to double H-bond structures, and the drastic reduction in the second peak, related to single H-bonded dimers.

The free energy of dimerization largely depends on the definition of "bonded" dimers. A dimerization free energy of 0.7 kcal/mol is estimated using a distance criterium. However, considering the population of H-bonded dimers, which seems more suitable to compare with IR-derived measures, a value of 2.4 kcal/mol is obtained. This value agrees with the experimental free energy of dimerization for cis amides (2.5 kcal/mol according to IR experiments),[2b,l] which supports our MC−MST calculations. There is a qualitative agreement between our results for formamide, and experimental measures for *trans*-amide dimerization in water,[2b] which suggests free energy differences of around 3.1 kcal/mol. There is also qualitative agreement with discrete MC−PMF calculations by Jorgensen,[3b] where free energies of dimerization ranging from 1.9 to 5.4 kcal/mol were determined depending on the cutoff radii used to discriminate between "bonded" and "non-bonded" configurations.

Figure 5. Areas of high probability to find a formamide molecule (for a reference space defined by the first formamide). The plots correspond to a density contour of 20 in gas phase (bottom) and chloroform (middle) and 5 in water (top).

Conclusions

High level ab initio calculations combined with ab initio SCRF calculations allowed us to describe accurately dimerization processes when there is a clear unique conformation for the bound dimer but provide incomplete representations in cases where there are a large multiplicity of bonded forms. In this case, molecular dynamics or Monte Carlo techniques should be used to obtain reasonable results.

Chloroform and especially water have a dramatic effect in the dimerization of formamide. The increase in the polarity of the solvent not only greatly reduces the dimerization but also has a profound influence on the nature of the "bonded" forms from well defined double H-bonded structures (gas phase) to single H-bonded dimers (chloroform) and non-well-defined structures (water). This dramatic change warns against the use of standard thermodynamic cycles for the computation of free energies of dimerization in solution.

The classical (solute−solute)−semiclassical (solvation) procedure followed in MC−MST calculations seems able to reproduce both formamide−formamide interaction energy in the gas phase and the solvation of formamide monomer and dimer. MC−MST calculations are able to provide an accurate picture of the configurational space accessible to the formamide dimer. The free energies of dimerization in gas phase, chloroform, and water predicted by MC−MST calculations agree well with available experimental results. Finally, MC−MST calculations allow us to explain many apparently paradoxical experimental results reported in the literature. All of these evidences give confidence on the quality of MC−MST calculations to describe the dimerization process.

Acknowledgment. We thank Dr. Tomasi for providing us with his version of PCM, which was modified by us to perform QM−SCRF calculations. This work was suported by the Centre de Supercomputació de Catalunya (CESCA, Mol. Recog. Project), as well by the Spanish DGICYT (PB96-1005 and PB97-0908). This is a contribution of the Centre Especial de Recerca en Química Teòrica de la Universitat de Barcelona.

Supporting Information Available: Table listing force-field parameters used to describe formamide−formamide interactions in MC−MST calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) (a) Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davis, D. R.; Philips, D. C.; Shore, V. G. *Nature* **1960**, *185*, 422. (b) Schultz, G. E.; Schrimer, R. H. *Principles of Protein Structure*; Springer-Verlag: New York, 1979.

(2) (a) Huisgen, R.; Walz, H. *Chem. Ber.* **1956**, *89*, 2616. (b) Klotz, I. M.; Franzen, J. S. *J. Am. Chem. Soc.* **1962**, *84*, 3461. (c) Susi, H.; Timasheff, S. N.; Ard, J. S. *J. Biol. Chem.* **1964**, *239*, 3051. (d) Affsprung, H. E.; Christian, S. D.; Worley, J. D. *Spectrochim. Acta* **1964**, *20*, 1415. (e) Susi, H. *J. Phys. Chem.* **1965**, *69*, 2799. (f) Hopmann, R. F. W. *J. Phys. Chem.* **1974**, *78*, 2341. (g) Wagner, K.; Rudakoff, G.; Frölich, P. *Z. Chem.* **1975**, *15*, 272. (h) Davis, M.; Thomas, D. K. *J. Phys. Chem.* **1975**, *79*, 767. (i) Walmsley, J. A. *J. Phys. Chem.* **1978**, *82*, 2031. (j) Josefiak, C.; Schneider, G. M. *J. Phys. Chem.* **1979**, *83*, 2126. (k) Spencer, J. N.; Garrett, R. C.; Mayer, F. J.; Merkle, J. E.; Powell, C. R.; Tran, M. T.; Berger, S. K. *Can. J. Chem.* **1980**, *58*, 1372. (l) Krikorian, S. E. *J. Phys. Chem.* **1982**, *86*, 1875. (m) Searle, M. S.; Williams, D. H.; Gerhard, U. *J. Am. Chem. Soc.* **1992**, *114*, 10697.

(3) (a) Hobza, P.; Mulder, F.; Sandorfy, C. *J. Am. Chem. Soc.* **1982**, *104*, 926. (b) Jorgensen, W. L. *J. Am. Chem. Soc.* **1989**, *111*, 3770. (c) Guo, H.; Karplus, M. *J. Phys. Chem.* **1992**, *96*, 7273. (d) Neuheuser, T.; Hess, B. A.; Reutel, C.; Weber, E. *J. Phys. Chem.* **1994**, *98*, 6459. (e) Dixon, D. A.; Dobbs, K. D. *J. Phys. Chem.* **1994**, *98*, 13435. (f) Florian, J.; Johnson, B. G. *J. Phys. Chem.* **1995**, *99*, 5899. (g) Suhai, S. *J. Chem. Phys.* **1995**, *103*, 7030. (h) Hobza, P.; Sponer, J. *J. Mol. Struct. (THEOCHEM)* **1996**, *388*, 115. (i) Tam, C. N.; Bour, P.; Eckert, J.; Trouw, F. R. *J. Phys. Chem. A* **1997**, *101*, 5877. (j) Sponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 263. (k) Ben-Tal, N.; Sitkoff, D.; Topol, I. A.; Yang, A.-S.; Burt, S. K.; Honig, B. *J. Phys. Chem.* **1997**, *101*, 450.

(4) Colominas, C.; Teixidó, J.; Cemelí, J.; Luque, F. J.; Orozco, M. *J. Phys. Chem.* **1997**, *102*, 22269.

(5) (a) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221. (b) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622. (c) Curtiss, L. A.; Jones, C.; Trucks, K.; Raghavachari, K.; Pople, J. A. *J. Chem.*

Phys. **1990**, *93*, 2537. (d) Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys*. **1993**, *98*, 1293.

(6) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(7) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Peterson, G. A.; Montgomery, G. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, Y. C.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; D. J., Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94* (Rev. D.3); Gaussian Inc.: Pittsburgh, PA, 1995.

(8) Jorgensen, W. L. BOSS3.4. Department of Chemistry, Yale University, 1993.

(9) (a) Momany, F. A. *J. Phys. Chem.* **1978**, *82*, 592. (b) Bonaccorsi, R.; Petrongolo, C.; Scrocco, E.; Tomasi, J. *Theor. Chim. Acta* **1971**, *20*, 331. (c) Orozco, M.; Luque, F. J. *J. Comput. Chem*. **1990**, *11*, 909. (c) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431. (d) Orozco, M.; Luque, F. J. *J. Comput. Aided Mol. Des*. **1990**, *4*, 411.

(10) (a) Becke, A. D. *J. Chem. Phys*. **1993**, *98*, 5648. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(11) (a) Soliva, R.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **1997**, *18*, 980. (b) Soliva, R.; Luque, F. J.; Orozco, M. *Theor. Chem. Acc*. **1997**, *98*, 42.

(12) (a) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117. (b) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.

(13) Luque, F. J.; Orozco, M. *J. Phys. Chem.* **1997**, *101*, 5573.

(14) (a) Orozco, M.; Luque, F. J. *Chem. Phys.* **1994**, *182*, 237. (b) Luque, F. J.; Bachs, M.; Orozco, M. *J. Comput. Chem.* **1994**, *15*, 847. (c) Orozco, M.; Bachs, M.; Luque, F. J. *J. Comput. Chem.* **1995**, *16*, 563. (d) Luque, F. J.; Bachs, M.; Alemán, C.; Orozco, M. *J. Comput. Chem.* **1996**, *17*, 806. (e) Luque, F. J.; Zhang, Y.; Alemán, C.; Bachs, M.; Gao, J.; Orozco, M. *J. Phys. Chem.* **1996**, *100*, 4269. (f) Luque, F. J.; Orozco, M. *J. Phys. Chem.* **1997**, *101*, 5573.

(15) (a) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717. (b) Claverie, P. In *Intermolecular Interactions: from Diatomics to Biomolecules*; Pullman, B., Ed.; Wiley: Chichester, 1978. (c) Colominas, C.; Luque, F. J.; Teixidó, J.; Orozco, M. *Chem. Phys*. **1999**, *240*, 253.

(16) (a) Luque, F. J.; Bofill, J. M.; Orozco, M. *J. Chem. Phys.* **1995**, *103*, 10183. (b) Ángyán, J. G. *J. Chem. Phys.* **1997**, *107*, 1291. (c) Luque, F. J.; Bofill, J. M.; Orozco, M. *J. Chem. Phys.* **1997**, *107*, 1293.

(17) Colominas, C.; Luque, F. J.; Orozco, M. *J. Comput. Chem*. **1999**, *20*, 665.

(18) Luque, F. J.; Alhambra, C.; Orozco, M. MOPETE/MOPFIT computer programs, University of Barcelona, Barcelona, 1998.

(19) Peterson, M.; Poirier, R. *MonsterGauss*. Department of Chemistry, University of Toronto, Canada; version modified by Cammi, R.; Bonaccorsi, R.; Tomasi, J. University of Pisa, Pisa, 1987; version modified by Luque, F. J.; Orozco, M. University of Barcelona, Barcelona, 1998.

(20) Colominas, C.; Luque, F. J.; Orozco, M. MC−MST computer program, University of Barcelona, Barcelona, 1998.