Structure of Liquid Formic Acid Investigated by First Principle and Classical Molecular Dynamics Simulations

Riccardo Chelli,*,†,‡,§ Roberto Righini,†,‡,§ and Salvatore Califano†,‡

Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, European Laboratory for Nonlinear Spectroscopy (LENS), Via Nello Carrara 1, 50019 Sesto Fiorentino, Italy, and INSTM-UdR Firenze, Italy

Received: April 5, 2005; In Final Form: June 17, 2005

The structure of liquid formic acid has been investigated by Car−Parrinello and classical molecular dynamics simulations, focusing on the characterization of the H-bond network and on the mutual arrangement of pairs of bonded molecules. In agreement with previous computational studies, two levels of H-bonded structures have been found. Small clusters, characterized by O−H···O bonds, are held together by weak C−H···O bonds to form large branched structures. From the ab initio simulation we infer the importance of cyclic H-bond dimer configurations, typical of the gas phase. Most of these dimer structures are however found to be embedded into H-bonded chains. When only O−H···O bonds are taken into account, linear H-bond chains are detected as basic structures of the liquid. More branched structures occur when C−H···O bonds are also considered. Regarding the arrangement of molecular pairs, we observed that O−H···O bonds favor the occurrence of configurations with parallel molecular planes, whereas no preferential orientation is observed for molecules forming C−H···O bonds.

1. Introduction

Formic acid is one of the simplest organic molecules and has been often considered as a model system of the carboxylic acid series. The molecule has C_s symmetry, and its flexibility is limited to the torsional motion of the hydroxy group around the C−OH bond. Several chemicophysical properties of formic acid depend on its capability to form different types of hydrogen bonds (H-bond). A formic acid molecule can actually act both as hydrogen-acceptor with the two oxygen atoms, and as hydrogen-donor with the hydroxy and formyl hydrogen atoms. Therefore four types of H-bonds are in principle possible: O−H···O_c, O−H···O_h, C−H···O_c, and C−H···O_h, where O_c and O_h are the carbonyl and hydroxy oxygen atoms, respectively. As a consequence, despite the great molecular simplicity, its aggregation states show a variety of different structures. In the gas phase[1−3] formic acid forms cyclic H-bonded dimers, whereas the structure of the crystal is characterized by infinite H-bonded chains, with alternate O−H···O and C−H···O bonds between the monomeric units.[4,5] Liquid formic acid is instead considered to exist in the form of a H-bonded network, involving a variety of chain like structures, different from both that of the gas and of crystal.[6−11] Moreover, a cis molecular structure, different from the trans structure existing in normal conditions (the trans/cis nomenclature refers to the position of the hydrogen atoms), has been observed in crystalline formic acid under high-pressure.[12]

Liquid formic acid has been investigated by a number of experimental techniques, including Raman[8,13,14] and NMR[15,16] spectroscopy, X-ray[6,9,10,17,18] and neutron[6,10,19] diffraction, and dielectric measurements.[20] The available experimental data on the liquid structure were essentially interpreted using relatively simple models.[6,15] Different conclusions were, however, reached on the structure of the liquid by various authors. In this respect computer simulations based on full-atom models can offer, in principle, a significant contribution to the understanding of the liquid structure at an atomic level because atomic trajectories are directly obtained from such methods. The structure and the dynamics of liquid formic acid related to the H-bonding have been indeed studied with computational techniques. A reverse Monte Carlo (MC) simulation, based on the simultaneous fit of X-ray and neutron diffraction structure functions, was carried out by Jedlovszky et al.[10] Furthermore, Jedlovszky, Turi et al. performed MC[11,21] and molecular dynamics[22] (MD) simulations, using a potential model tuned on the ab initio potential energy surface of formic acid dimers.[21] Hermida Ramón and Ríos[23] have developed a polarizable force field for *trans*-formic acid, devised for MD simulations in liquid phase. Their analysis was focused on the influence of the induction mechanisms on chemicophysical properties such as the molecular dipole moment and the pair interaction energy of the molecules.

In the framework of a research project on the dynamics of H-bonded liquids, we have studied by optical Kerr effect the short time decay processes of the molecular motions in liquid formic acid (data not published). To interpret such experiments at the microscopic level, we are planning to use MD simulations, as done previously for other systems.[24] For a correct simulation of the dynamics it is, however, necessary to be sure that the intermolecular potential model correctly reproduces the stucture of the liquid. For this reason we present here an ab initio and a classical MD simulation aimed to a description of the structural properties of the liquid formic acid. In a forthcoming paper, using the present empirical model, we shall address the dynamical properties of the system.

The ab initio simulation is based on the first principle approach developed by Car and Parrinello,[25] while the classical

* Corresponding author. E-mail: chelli@chim.unifi.it. Fax: 0039 0554573077; Phone: 0039 0554573082.
† Università di Firenze.
‡ European Laboratory for Nonlinear Spectroscopy.
§ INSTM-UdR.

10.1021/jp051731u CCC: $30.25
Published on Web 08/13/2005

simulation is based on an empirical potential model[26] originally devised for biochemical systems. The Car−Parrinello simulation has the advantage of avoiding the use of an empirical potential, thus offering a better description of the intra- and intermolecular interactions in the liquid. On the other hand, owing to the very large computational cost, the dimension of the simulation box is necessarily small and the simulation time is short as well. Therefore a statistical mechanic approach to the liquid structure is less significant. The opposite situation occurs for the classical simulation. The dimension of the simulation box can be reasonably larger to extend the radius of the atom−atom interactions at least to the third coordination shell, for a more sound statistical approach. The use of an empirical intermolecular potential can, however, neglect some interactions of quantum nature such as those involving the occurrence of resonant structures. Comparison between the results of the two simulation methods provides therefore an important protocol to reduce the model bias and hence to better validate the results. The ab initio simulation is used to control whether the empirical potential used in the classical simulation is accurate enough. On the other hand the classical simulation is utilized to check the occurrence of possible box size effects in the ab initio data and to furnish a more significant statistics. In the present study we analyze important structural properties of the liquid formic acid at an atomic level. Special attention is devoted to the understanding of the role of H-bonding in the pair molecular arrangement. Furthermore the H-bond network structure of the liquid is investigated along with the role played by cyclic H-bonded dimers.

The article is organized as follows. The computational details are described in section 2, and the results are reported and discussed in section 3. Conclusions are given in section 4.

2. Simulation Methods

2.1. Car−Parrinello Molecular Dynamics Simulation. The Car−Parrinello MD simulation (CP) was made on 26 molecules of deuterated formic acid placed in a cubic box. As starting geometry, the molecules were assumed in the trans conformation. Given the large energy difference between the two possible conformers[27] (16.3 kJ mol^{-1}), the trans conformer is expected to be dominant in both the liquid and the gas phases. As a matter of fact this assumption is the one usually accepted in the early theoretical and experimental studies on liquid formic acid (see, e.g., refs 15, 19, 21, and 23). However the simulations were performed without constraints on the torsion angles, and thus conformational transitions were not forbidden a priori. Despite this rotational freedom, conformational changes were not observed neither in the ab initio nor in the classical simulation. The sample was prepared by first performing a constant particle, volume, and energy simulation (NVE) using the empirical potential devised for the classical MD simulation (CS) described in section 2.2. The volume of the simulation box was chosen to reproduce the experimental value[28] of 0.01588 molecules per Å^3. The system was equilibrated at 298 K by a run of 100 ps with scaling of the atomic velocities, followed by a run of 200 ps without velocity scaling. The last system configuration obtained from the CS simulation was then used as starting geometry for the CP simulation, made in the NVE thermodynamical ensemble. Standard periodic boundary conditions were applied. During the first 1.2 ps of the CP simulation, the scaling of the atomic velocities was active. During the production run, of about 11.8 ps, 19574 system configurations were stored.

Martins−Troullier pseudopotentials[29] were used along with the Kleinman−Bylander decomposition.[30] The plane wave expansions were truncated at 70 Ry. Calculations were performed in the generalized gradient approximation using the BLYP exchange-correlation functional.[31,32] The equation of motions were integrated with a time step of 5 au (∼0.12 fs). Deuterium was employed instead of hydrogen to allow for a larger time-step, while a fictitious electronic mass of 800 au was adopted to keep the system on the Born−Oppenheimer surface. The simulation was performed with the CPMD program.[33]

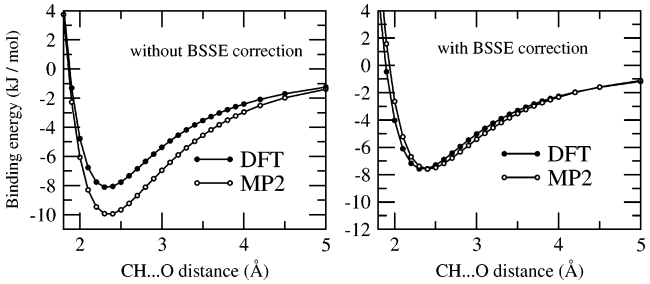

Figure 1. Binding energy profile of a formic acid dimer with a C−H···O interaction established (see text for geometrical details on the dimer). B3LYP/6-311++G** (full circles) and MP2/6-311++G** (open circles) calculations are reported with and without including BSSE correction.

In a recent paper Wang et al.[34] raised an important question concerning the ability of density functional theory (DFT) methods to reproduce the C−H···O interaction. In particular they calculated the binding energy of 1:1 complexes of *N*-methyl maleimide (NMM) with dimethyl sulfoxide, acetone, and water, using configurations where bifurcated C−H···O bonds occur. The conclusion of Wang et al. was that DFT (with B3LYP exchange-correlation functional) significantly underestimates the binding energy with respect to the second-order Møller−Plesset (MP2) perturbation theory.[35] The authors ascribed this underestimate to the inability of DFT to reproduce the dispersion component of the bonding. This is a longtime known problem of DFT methods,[36] and given the importance of the C−H···O interactions in the present study, it is essential to assess the validity of the DFT approach in the specific case of formic acid. Even in the light of the results by Wang et al.,[34] this issue is not obvious. In fact, because of the different chemical environment experienced by the CH group in formic acid with respect to NMM (at variance with NMM, the CH group of formic acid is directly bonded to two strong electron-withdrawing groups), the behavior of DFT and MP2 may also be remarkably different in the two cases.

We estimated the binding energy profile for a formic acid dimer in which a C−H···O bond occurs. To allow the comparison with the calculations of Wang et al.,[34] we used the B3LYP and the MP2 ab initio methods along with the 6-311++G** basis set. Fixed geometry and mutual orientation of the molecules were assumed. In particular the molecules were kept perpendicular each other with the four atoms C−H···O═C placed on the same axis. With these constraints, the dimeric geometry is unambiguously established. The single point ab initio calculations were carried out by varying the CH···O distance, and the binding energy profile was estimated with and without including the basis set superposition error (BSSE) accounted for by the counterpoise method.[37] The binding energy profile is shown in Figure 1 before and after BSSE correction. When BSSE is accounted for, B3LYP and MP2 give essentially the same profile, the only detectable difference being a small shift of the curves. In the case of the NMM complexes,[34] the difference between the B3LYP and MP2 binding energies (including BSSE correction) is 2.5, 1.8, and 1.3 kJ mol^{-1} for

TABLE 1: Parameters of the Intermolecular Potential and Molecular Geometry[a]

atom type	σ (Å)	ϵ (kJ mol^{-1})	q (electrons)
C	3.4000	0.359824	0.543715
O_c	2.9596	0.878640	−0.473810
O_h	3.0665	0.880314	−0.546537
H_f	2.2932	0.065689	0.048715
H_h	0.0000	0.000000	0.427917

molecular equilibrium geometry

bond (Å)		angle (deg)	
$C-O_c$	1.2111	$C-O_h-H_h$	107.33
$C-O_h$	1.3658	H_f-C-O_c	125.63
O_h-H_h	0.9818	O_c-C-O_h	125.33
$C-H_f$	1.1047	$O_c-C-O_h-H_h$	0.00

[a] The Lennard−Jones parameters refer to the potential form $V = 4\epsilon_{ij}[(\sigma_{ij}/R_{ij})^{12} - (\sigma_{ij}/R_{ij})^6]$ where $\epsilon_{ij} = (\epsilon_i\epsilon_j)^{1/2}$ and $\sigma_{ij} = (\sigma_i + \sigma_j)/2$. Atom types have the following meaning: C is the carbon, O_c is the carbonyl oxygen, O_h is the hydroxy oxygen, H_h is the hydroxy hydrogen, and H_f is the formyl hydrogen.

dimethyl sulfoxide, acetone, and water, respectively. Therefore our results suggest that in formic acid the dispersion component does not affect the C−H···O interaction significantly. Finally we stress that, because of the use of plane waves, our CP simulation is BSSE free.

2.2. Classical Molecular Dynamics Simulation. The CS simulation was performed for a system of 500 deuterated molecules. The potential model is based on the full-atom parametrization proposed by Cornell et al.[26] that treats the intermolecular interactions by Lennard-Jones and Coulombic pair interactions. The point charges on the atoms were calculated by a fit of the electrostatic potential around the molecule at the B3LYP/6-31G(d) level of theory on a molecular geometry optimized at the B3LYP/6-311++G(d,p) level. The calculation of the atomic charges was carried out using the Gaussian 98 program.[38] The Lennard−Jones parameters and the atomic charges are listed in Table 1. The force constants of ref 26 were used for torsions and for bending angles, while the covalent bonds were constrained using the RATTLE algorithm.[39] The equilibrium values of the bond distances and bending angles were determined by geometry optimization of the molecule at the B3LYP/6-311++G(d,p) level of theory. The equilibrium geometry of the formic acid molecule is reported in Table 1.

The starting configuration was built by distributing the carbon atoms at the nodes of a cubic grid in the simulation box with random orientations of the molecules. The system was equilibrated for a short period (20 ps) at 400 K under isobaric−isothermal thermodynamic conditions (NPT ensemble using the isotropic stress tensor constraint). The temperature was kept constant by a Nosé−Hoover thermostat,[40,41] while constant pressure was imposed using a modification of the Parrinello−Rahman Lagrangian.[42] A NPT equilibration run at 298 K with scaling of the atomic velocities followed for about 100 ps. A further MD simulation for equilibrating the system was carried out for 200 ps without scaling of the atomic velocities. Finally a simulation run of 640 ps was performed to calculate the average volume of the simulation box. This last run was used also to compute some thermodynamical properties of the system (see section 3.1). Using the average volume and the last system configuration obtained from the NPT simulation, the system was then equilibrated (200 ps with atomic velocity scaling and 500 ps without atomic velocity scaling) with NVE simulations imposing a temperature of 298 K. After equilibration, a NVE production run of 1.2 ns was carried out to record the atomic trajectory (3×10^4 system configurations).

TABLE 2: Thermodynamical Properties and Self-diffusion Coefficient of Liquid Formic Acid

	ρ g cm^{-3}	ΔH_{vap} kJ mol^{-1}	C_p J mol^{-1} K^{-1}	β_T, 10^{-10} Pa^{-1}	D 10^{-5} cm^2 s^{-1}
CS simul	1.213	31.7[d]−40.1[e]	100.06	5.08	1.44
exp	1.214[a]	38.96[b]	99.04[a]	6.47[a]	1.04[c]

[a] From ref 28; [b] From ref 46; [c] From ref 47; [d] Calculated assuming a gas of isolated cyclic dymers; [e] Calculated assuming a gas of isolated monomers.

Coulombic interactions were treated by the smooth particle mesh Ewald method[43] using a convergence parameter of 0.43 Å^{-1} and a grid spacing for the B-spline interpolation of charges of about 0.9 Å. The multiple time step r-RESPA technique[44] was employed to integrate the equations of motion. The cutoff radius for the direct lattice electrostatic and Lennard−Jones interactions is 11.5 Å. Standard periodic boundary conditions were applied. The simulations were performed using the ORAC program.[45]

3. Results and Discussion

3.1. Thermodynamical and Dynamical Properties. Some thermodynamical properties and the self-diffusion coefficient of liquid formic acid calculated from the CS simulation are listed in Table 2 along with the experimental data.[28,46,47] The density ρ calculated from the NPT simulation agrees very well with the experimental value.[28] Since the gas composition is not known, the vaporization enthalpy ΔH_{vap} at 298 K was computed for two limiting cases: (i) a gas formed by isolated monomers alone; (ii) a gas formed by isolated cyclic dimers[1−3] alone. The real composition of the gas is probably a mixture of the two species and perhaps of other less abundant ones. The experimental vaporization enthalpy should thus fall between the two calculated limiting values. The following equation was used for both the cases:

$$\Delta H_{vap} = E^{gas}_{inter} + E^{gas}_{intra} - E^{liq}_{inter} - E^{liq}_{intra} + nRT \qquad (1)$$

The $E^{liq}_{inter} + E^{liq}_{intra}$ quantity is independent of the assumption of the gas composition. Its value (−128.94 kJ mol^{-1}) was taken from the NPT run described in section 2.2. E^{gas}_{intra} for the monomer-formed gas (−91.32 kJ mol^{-1}) was determined from an average of multiple MD simulations of the monomer, while E^{gas}_{inter} was neglected. In the dimer-gas approximation, the full quantity $E^{gas}_{inter} + E^{gas}_{intra}$ was calculated from an average of multiple MD simulations of the cyclic dimer (−98.51 kJ mol^{-1}). Each simulation was stopped at the occurrence of the dimer break down (in average after 100−300 ps). Finally, the value of n in eq 1 is assumed to be 1 and 0.5 for the monomer and dimer-gas approximations, respectively. The experimental ΔH_{vap} is in reasonable agreement with both calculated values (see Table 2). Due to the small difference between the two calculated enthalpies, we do not trust, however, to estimate the gas composition from these results. The molar specific heat at constant pressure C_p, calculated from the enthalpy fluctuations,[48] agrees within 1% with the experimental value.[28] The isothermal compressibility β_T was computed from the volume fluctuations in the NPT ensemble.[48] Although the difference from the experimental value[28] exceeds 20%, the result is slightly better than that of the earlier simulation by Jedlovszky and Turi.[21] As pointed out in ref 21, this is due to the slow convergence of the simulation box volume. Another important quantity for assessing the correctness of the model is the self-diffusion coefficient. Using the Einstein relation[48] we obtained a coefficient of 1.44

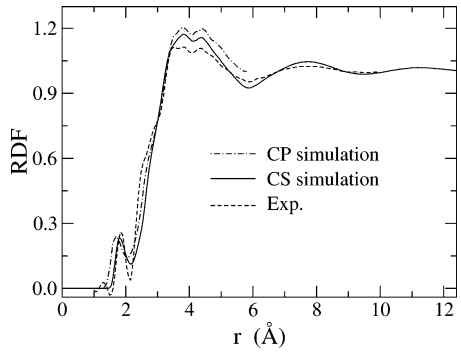

Figure 2. Calculated and experimental radial distribution functions of liquid formic acid. Experimental data from ref 10.

$\times 10^{-5}$ cm^2 s^{-1}, in reasonable agreement with the experimental value[47] of 1.04 $\times 10^{-5}$ cm^2 s^{-1}. A further dynamical property that can be compared to an experimental counterpart is the time decay of the single particle autocorrelation function of the first Legendre polynomial calculated from the molecular dipole moment, *i.e.*, $\Sigma_{i=1}^{N} \langle \boldsymbol{\mu}_i(t) \cdot \boldsymbol{\mu}_i(0) \rangle$. This quantity can be easily related to dielectric properties[22] of the system, and, since it originates from an average of single-molecule functions, it converges much faster than collective quantities such as the dielectric constant. From the autocorrelation function we obtained a fitted time decay of about 6.6 ps, while the integral correlation time is 5.5 ps. While these estimates are comparable to the values of 6.0 and 5.0 ps obtained by Mináry et al. in a previous MD simulation,[22] they are quite different from the single-particle correlation time (3.1 ps) estimated in ref 22 on the basis of dielectric measurements.[20]

The results reported in this section show that the adopted empirical model is sufficiently accurate to perform a realistic CS simulation on formic acid. As we shall see in the following sections, the reliability of the model is further confirmed by the comparison of several structural properties with those calculated from the CP simulation.

3.2. Radial Distribution Functions. The calculated radial distribution functions (see ref 48 for computational details), from now on indicated as RDF, are compared with neutron scattering data[10] in Figure 2. The agreement with experiment is satisfactory for both CS and CP simulations. The CS simulation reproduces well the experimental H-bond peak at about 1.85 Å, whereas the CP one predicts slightly stronger H-bond interactions, the peak falling at about 1.7 Å. In addition even the oscillations of the experimental curve up to 10 Å are present in the CS curve, although with a larger amplitude.[21,49] A further difference between calculated and experimental RDFs concerns the shoulder occurring at about 2.7 Å in the neutron scattering measurements,[6,10] which is less evident in the simulated curves (especially in the CS one) although still visible. Given the short distance, the origin of this shoulder is attributed to structural features of H-bonded molecules.[6]

For a correct assignment of the observed peaks to the different pair interactions, we show in Figure 3 the most significant calculated partial RDFs. Only the three RDFs involving pairs of hydrogen atoms are not shown in the figure since, due to the small scattering amplitude of deuterium with respect to oxygen and carbon, their contribution to the total RDF is negligible. The partial RDFs globally agree with those obtained from a reverse MC simulation.[10] The experimental peak at 1.85 Å is clearly assigned to O−H···O$_c$ bonds as proved by the sharp and intense first peak of the $g_{O_cH_h}(r)$ falling at 1.8 Å for the CS and 1.7 Å for the CP simulation. The O−H···O$_h$ bonds are instead much weaker than the O−H···O$_c$ ones. In fact the peak

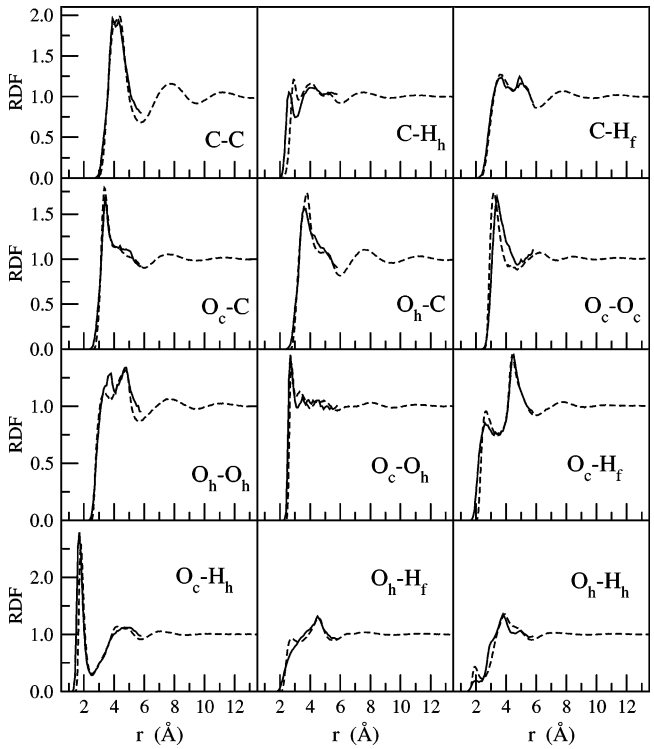

Figure 3. Partial radial distribution functions of liquid formic acid for some atomic pairs. Solid line: CP simulation; dashed line: CS simulation.

associated to the H-bond in the $g_{O_hH_h}(r)$ not only has a lower intensity but also falls at a larger distance (i.e. about 2 Å) with respect to that of the main peak in the $g_{O_cH_h}(r)$.

As stated before, the shoulder at about 2.7 Å in the neutron scattering measurements of Figure 2 is associated to the H-bonded molecules. Only the $g_{O_cO_h}(r)$, $g_{CH_h}(r)$, $g_{O_cH_f}(r)$, and $g_{O_hH_f}(r)$ RDFs significantly contribute in the 2.6−2.8 Å range. Owing to the small scattering amplitude of deuterium with respect to oxygen and carbon, the contribution from $g_{CH_h}(r)$, $g_{O_cH_f}(r)$, and $g_{O_hH_f}(r)$ is small and thus the peak intensity is dominated by the contribution of the $g_{O_cO_h}(r)$ RDF.

The CP $g_{O_hO_h}(r)$ RDF shows a peak at 3.7 Å, occurring also in the reverse MC calculations.[10] This distance, as well as the shoulder at about 3 Å in the $g_{O_hH_h}(r)$ RDF, is a typical signature of the occurrence of cyclic H-bond dimer configurations with two O−H···O$_c$ bonds.

An important characteristic of formic acid is its ability to form nonconventional C−H···O bonds. These H-bonds were found in the crystal,[4,5] along with those of O−H···O type. There are both experimental[15] and theoretical[11] indications of the occurrence of C−H···O bonds also in the liquid. The $g_{O_hH_f}(r)$ and $g_{O_cH_f}(r)$ RDFs confirm such indications, showing shoulders in the 2.0−2.5 Å range that can be explained only assuming significant C−H···O interactions between the molecules. From the RDFs we infer that the C−H···O$_c$ bonds are more stable than the C−H···O$_h$ ones, in line with the larger stability of the O−H···O$_c$ bonds with respect to the O−H···O$_h$ ones. This is quantitatively proved by the coordination numbers calculated from $g_{O_cH_f}(r)$ and $g_{O_hH_f}(r)$ by integration up to 2.5 Å, which are 0.37 and 0.19, respectively.

3.3. H-Bond Statistics and H-Bond Dimer Configurations. To investigate the H-bond formation in liquid formic acid we have used a geometrical criterion for the H-bond definition.[50−52] To obtain comparable results for the O−H···O and C−H···O bonds, a common criterion has been assumed. We consider a

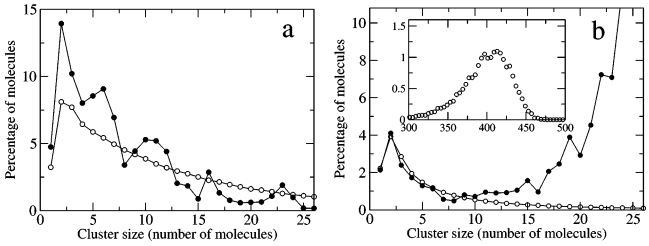

Figure 4. Population distribution of formic acid molecules as a function of the H-bond cluster size. Graph a: distribution considering only O−H···O bonds. Graph b: distribution considering both O−H···O and C−H···O bonds. Open circles: CS simulation; filled circles: CP simulation. The solid lines are drawn as guides to the eyes.

X−H···O bond to be established when the X···O distance is less than 2.5 Å (the first minimum in the $g_{O_cH_h}(r)$ RDF) and the X−H···O angle greater than 130°. In addition this choice allows us to compare our results with those obtained by Jedlovszky and Turi from a MC simulation.[11]

Several approaches, based on the H-bond geometry, have been used in previous works[10,11] to study the connectivity of H-bonded molecules in formic acid. The basic information emerging from these works is the occurrence of small clusters with O−H···O intermolecular bonds, connected each to the other by C−H···O bonds to form a space-filling network. To characterize the H-bond network, we have first evaluated the size and the abundance of the O−H···O bonded domains in the liquid and then their association through C−H···O bonds. In these calculations a collection of molecules is considered to form a cluster if every pair of molecules is connected by a chain of O−H···O bonds. In the same way two clusters are considered to form a larger cluster, if at least one molecule of the first and one of the second are connected through a C−H···O bond. The molecular population distribution as a function of the dimension of the clusters has been calculated using both the CP and the CS simulations. The results obtained without and with inclusion of the C−H···O bonds are shown in Figures 4a and 4b, respectively. Figure 4a shows that the percentage of molecules involved in O−H···O clustering decreases as the cluster size increases. Owing to the small dimension of the simulation box (and thus to the low statistical sampling), the curve obtained from the CP simulation is more noisy than that calculated from the CS one. However the trend of the two curves is essentially the same. When the C−H···O bonds are included in the calculation (Figure 4b), the population associated to small size clusters decreases because of the formation of larger clusters. Most of the population moves toward very large H-bonded cluster sizes, as shown by the inset of Figure 4b. The curve obtained from the CS simulation decreases with the cluster size, reaches the zero for clusters of about 25 molecules, and then starts to grow up again above a cluster size of about 300 molecules. By integrating the population of the CS simulation (Figure 4b) up to the limit of 25 molecules, we estimate that about 19% of the molecules is involved in small H-bonded clusters. As a consequence, about 81% of the molecules forms large size clusters. Because of the small box size (also for the CS simulation), we can only estimate a lower limit for the size of such large clusters (see inset of Figure 4b). Below the cluster size of 10 molecules the CS and the CP simulations agree almost perfectly, indicating that no box size effects occur. An evident increase of population is instead observed for the CP simulation curve above this limit, clearly originating from box size effects.

In Table 3 we report the number of different O−H···O and C−H···O bonds per molecule as well as their total. The results of the CS simulation agree basically with those of the CP one.

TABLE 3: Number of H-Bonds per Molecule in Liquid Formic Acid

H-bonds	CP simul	CS simul
$O-H\cdots O_c$	1.63	1.46
$O-H\cdots O_h$	0.18	0.34
$C-H\cdots O_c$	0.38	0.20
$C-H\cdots O_h$	0.18	0.14
Total O−H···O	1.81	1.80
Total C−H···O	0.56	0.34
Total	2.37	2.14

As expected, the largest contribution to the total number of H-bonds per molecule is given by the $O-H\cdots O_c$ bonds, the number of other H-bonds being much smaller. The total number of conventional O−H···O bonds (1.81 for the CP and 1.80 for the CS simulation) agrees with the value of 1.85 obtained in an earlier simulation.[11] A significant difference between the CP and the CS simulation lies in the number of C−H···O bonds, which is larger in the former simulation. Moreover the number of C−H···O bonds per molecule predicted by our simulations is about one-half of that calculated by Jedlovszky and Turi[11] which found one C−H···O bond per molecule. We believe that this last value is largely overestimated because of the different strenght between the C−H···O and the O−H···O bond.

Another result emerging from Table 3 is that the number of $O-H\cdots O_h$ bonds is comparable (CS simulation), or even lower (CP simulation), to that of the C−H···O bonds. This cannot be due to the different energy stability between the two types of H-bond, since ab initio calculations[53] on cyclic dimers have shown that the $O-H\cdots O_h$ bond is about four times more stable than the C−H···O one (12.6−14.2 kJ mol^{-1} for $O-H\cdots O_h$ against 2.8−3.3 kJ mol^{-1} for C−H···O). We suggest that the observed behavior results from entropic factors due to the different availability of the H_h and H_f hydrogen atoms for H-bonding. Actually most of the hydroxy hydrogens are involved in the strong $O-H\cdots O_c$ bonds, whereas the formyl hydrogens are completely available to interact with both O_h and O_c oxygen atoms.

An interesting structural feature of liquid formic acid concerns the existence of stable cyclic H-bonded dimers. Theoretical studies have shown that the cyclic dimer with two $O-H\cdots O_c$ bonds, that is the dominant form in the gas phase,[1−3] is energetically the most stable.[53] However, the occurrence of an appreciable number of cyclic dimers in the liquid phase is still a matter of discussion. Several authors have concluded that liquid formic acid contains dimers,[15,16,54−56] whereas others suggest that the liquid contains essentially polymeric species.[10,11,13,17,20,57] In some cases a mixture of chainlike and dimer structures has also been invoked to explain the experimental observations.[6−9] From a MC simulation the presence of cyclic H-bonded dimers was found to be negligible with respect to other types of structure.[11]

In principle, formic acid can assume several different dimer configurations[53] with both O−H···O and C−H···O bonds. A schematic picture of these configurations is shown in Figure 5. In the figure we add also the dimerization energies obtained with the present empirical model (by energy minimization with the conjugate gradient method[58]) along with the ab initio counterpart calculated by Turi using the MP2/D95++(d,p) level of theory.[53] A check of the low frequency vibrations for saddle points has shown that all calculated structures are actually local minimum energy configurations. The empirical dimerization energies agree well with those obtained from the ab initio calculations of ref 53, thus offering an additional confirmation of the validity of the simulation model.

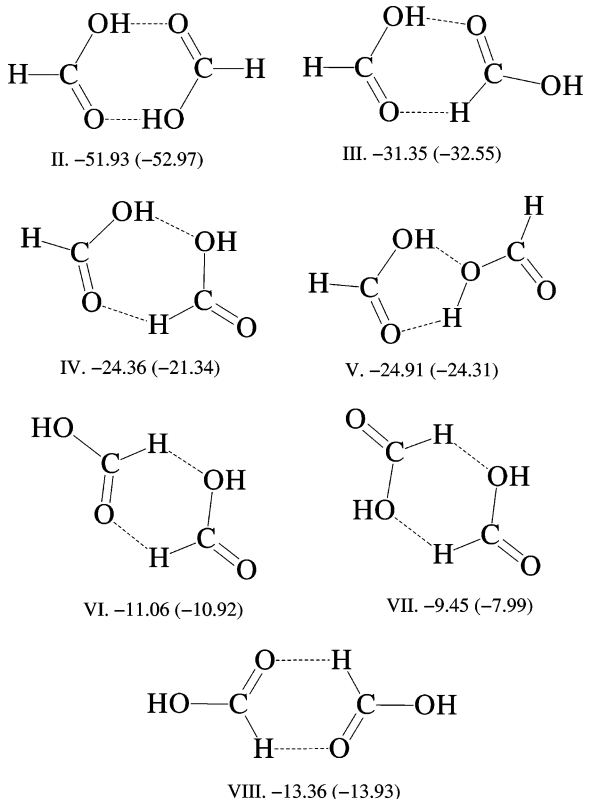

Figure 5. Schematic structures and dimerization energies (in kJ mol^{-1}) of cyclic dimers evaluated with the empirical model. Numbers in parentheses are the dimerization energies obtained from ab initio calculations.[53]

On the other hand, very accurate dimerization energies of the type II and type III dimers were reported recently by Brinkmann et al.[59] using various high level ab initio calculations. Using the explicitly correlated MP2 method of Kutzelnigg and Klopper,[60] they found dimerization energies of −67.4 and −40.8 kJ mol^{-1} for the type II and type III dimers, respectively. These values are larger than those calculated from our empirical model and the MP2/D95++(d,p) method[53] (see Figure 5). To assess the ability of the CP simulation to describe the interactions occurring in dimers, we calculated the dimerization energies of type II and type III dimers by the BLYP/plane waves method. We obtained −75.4 and −34.7 kJ mol^{-1} for the former and the latter type dimer, respectively. Both values are in satisfactory agreement with those of ref 59. In particular the dimerization energy of the type II dimer is remarkably higher than that calculated from our empirical model and from the model proposed by Jedlovszky and Turi.[21] We shall see below that this fact could be very insightful to explain the different concentration of the type II dimers predicted by the CP simulation and the simulations based on empirical models (including ours).

To gain further information on the structure of liquid formic acid in terms of competition between cyclic H-bonded dimers and polymeric chains, we have performed a statistical analysis of the fraction of molecules involved in dimerization, classifying the data on the basis of the occurrence of the various dimer structures of Figure 5. The results are given in Table 4, where the percentage of molecules forming cyclic H-bond dimer configurations is reported. The occurrence of cyclic dimers does not prevent, however, the formation of H-bonded chains, since the dimers, due to the large capability of a formic acid molecule to form several H-bonds, have the possibility of being embedded into the chains. The percentage of molecules forming isolated cyclic dimers, i.e., not bound to neighboring molecules, is also reported in Table 4. The relative population of the various types of dimer is fully consistent with the dimerization energies of Figure 5. Both simulations predict type II dimers as the most probable ones. The second most probable dimers, even if at a much lower extent, are those of type III, other dimers having a negligible probability of occurrence. Although the percentages of molecules forming isolated dimers of type II obtained from the two simulations are comparable (3.1 for the CP and 2.6 for the CS simulation), the CS simulation predicts a significantly lower number of molecules forming dimers of type II H-bonded to other molecules (20.4 for the CP and 6.8 for the CS simulation). As mentioned above, this last result should be hardly surprising because of the large difference between the dimerization energies of the type II dimer obtained from the empirical model and from the level of theory adopted in the CP simulation.

TABLE 4: Percentage of Molecules Involved in Cyclic H-Bond Dimerization

	CP simulation		CS simulation	
dimer type[a]	total[b]	isolated[c]	total[b]	isolated[c]
II	23.5	3.1	9.4	2.6
III	1.0	0.0	0.4	0.0
others[d]	0.8	0.0	0.2	0.0

[a] See Figure 5 for dimer labeling. [b] Total percentage of molecules involved in cyclic dimers. [c] Percentage of molecules involved in cyclic dimers not H-bound with the environment. [d] Total contribution of the cyclic dimers IV, V, VI, VII, and VIII.

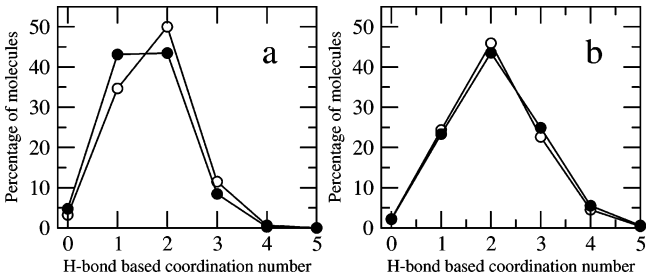

Figure 6. Population distribution of formic acid molecules as a function of the H-bond-based coordination number. Graph a: distribution considering only O−H···O bonds. Graph b: distribution considering both O−H···O and C−H···O bonds. Open circles: CS simulation; filled circles: CP simulation. The solid lines are drawn as guides to the eyes.

However the basic result of this analysis is that, in liquid formic acid, a significant number of molecules could indeed form stable cyclic H-bond dimers, typical of the gas phase. Most of these dimers are found to be embedded into extended H-bond structures.

A more exhaustive view of the H-bond molecular linkage in liquid formic acid can be obtained from an estimate of the H-bond based coordination number, defined as the average number of molecules bonded to a molecule with one or more H-bonds. In Figure 6a we report the population of the H-bond based coordination number considering only the O−H···O bonds. The analogous quantity calculated considering both O−H···O and C−H···O bond types is shown in Figure 6b. The agreement between the CS and CP simulation is very satisfactory in the case of Figure 6a and becomes almost perfect in the case of Figure 6b. When only O−H···O interactions are taken into account, the majority of molecules are H-bonded to one or two other molecules, suggesting that the linear H-bond chains are the most probable structures in liquid formic acid. From the data of Figures 4a and 6a, we also infer that about 30% of the molecules are located at the end position of H-bonded chains.

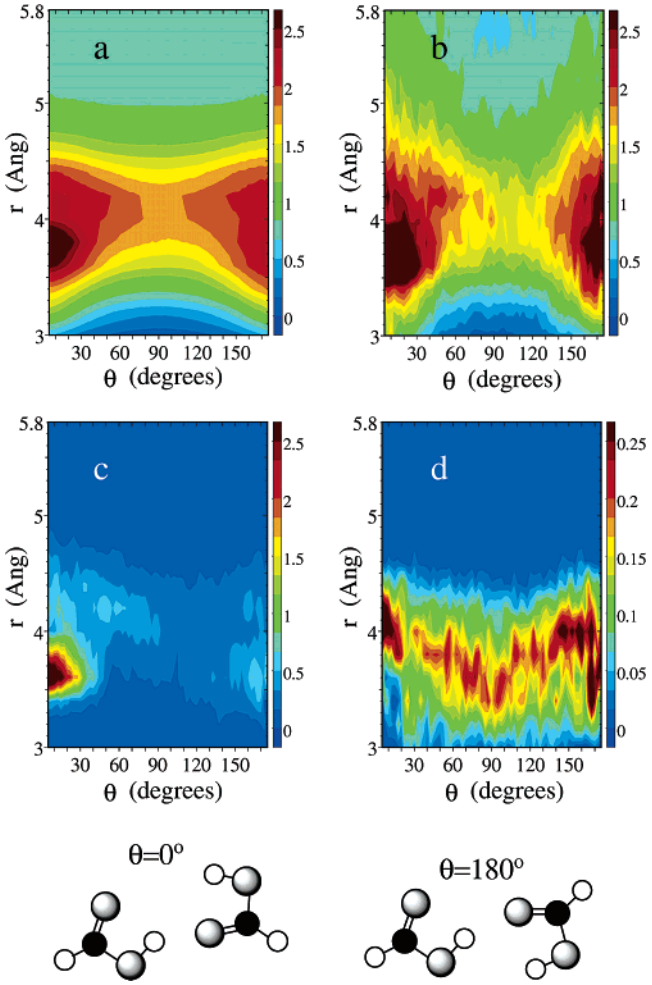

Figure 7. Radial−angular pair distribution functions, $G(r,\theta)$ (z-axis). Graph a: total distribution from the CS simulation. Graph b: total distribution from the CP simulation. Graph c: distribution from the CP simulation considering only pairs of molecules connected by O−H···O bonds. Graph d: distribution from the CP simulation considering only pairs of molecules connected by C−H···O bonds. The brown regions correspond to out of scale values, while the dark blue ones correspond to the zero value. Bottom: examples of pair configurations corresponding to $\theta = 0°$ and $\theta = 180°$.

If both types of H-bond (O−H···O and C−H···O) are considered, the H-bond based coordination number increases and more branched chains occur (see Figure 6b).

3.4. Role of H-Bonding in Pair Molecular Arrangement. In the previous section, some structural properties of liquid formic acid have been discussed in terms of H-bond linkage and of the topology of the molecular networking (H-bonded chains, cyclic dimer configurations, etc.). In this section we present instead an analysis aimed at determining the mutual arrangement of molecular pairs in space and the role of H-bonding in shaping the short-range structure of the liquid. The space arrangement of pairs of formic acid molecules has been analyzed in terms of the radial−angular pair distribution function[24,61,62], $G(r,\theta)$, defined as follows:

$$G(r,\theta) = \frac{V\langle \sum_{i=1}^{M}\sum_{j\neq i}^{M} \delta(r - R_{ij})\delta(\theta - \theta_{ij})\rangle}{2\pi M^2 r^2 \sin\theta \Delta r \Delta\theta} \qquad (2)$$

where R_{ij} is the distance between the centers of the molecules i and j, considered to be the carbon atoms, θ_{ij} is the angle formed by the normals to the molecular planes defined by the three heavy atoms, M is the number of molecules, V is the volume of the simulation box, Δr and $\Delta\theta$ are the computational resolutions for r and θ, respectively, $\delta(x)$ is the Dirac function, and the angular brackets indicate the time average. The global CS and CP computed $G(r,\theta)$ functions are presented in Figures 7a and 7b, respectively, and are very similar. This similarity is also evident in the radial−angular pair distribution functions limited to H-bonded molecules alone, i.e., $G_{O-H\cdots O}(r,\theta)$ and $G_{C-H\cdots O}(r,\theta)$ (data not shown for the CS simulation). These last quantities calculated from the CP simulation are reported in Figures 7c and 7d, respectively. The simulations predict the parallel arrangement of the molecules as the most probable one, with $\theta = 0°$ slightly more probable than $\theta = 180°$. For the sake of clarity, two schematic pair configurations corresponding to $\theta = 0°$ and $\theta = 180°$ are shown at the bottom of Figure 7. For $\theta = 0°$, the two molecules can overlap through a translation (in the three directions of the space) and a rotation around an axis perpendicular to the molecules. To overlap two molecules with $\theta = 180°$, an additional rotation around an axis in the molecular plane is required. A sharp and intense peak at about $\theta = 0°$ and $r = 3.6$ Å is evident in the $G_{O-H\cdots O}(r,\theta)$. A less pronounced peak occurs at $\theta \approx 180°$ and r ≈ 3.6 Å. From an analysis of the pair configurations as a function of r and θ, we estimate that the main peak originates from pair configurations of type II, while the smaller peak is mainly due to dimers of type III. In other regions of the (r,θ) space, the dominant contribution to $G_{O-H\cdots O}(r,\theta)$ originates from pairs of molecules forming only one O−H···O bond. This result is at variance with that of a previous MC simulation,[11] where no preferential arrangement of the planes of H-bonded molecules was observed.

In agreement with the trend observed in the partial RDFs discussed in section 3.2, the function $G_{C-H\cdots O}(r,\theta)$ gives a much smaller contribution to the total $G(r,\theta)$ with respect to $G_{O-H\cdots O}(r,\theta)$ (notice the different chromatic scale used in Figure 7 for $G_{O-H\cdots O}(r,\theta)$ and $G_{C-H\cdots O}(r,\theta)$). A remarkable feature of the $G_{C-H\cdots O}(r,\theta)$ function is its almost constant trend as a function of θ, in contrast to the $G_{O-H\cdots O}(r,\theta)$ which is strongly peaked at $\theta = 0°$. This indicates that molecules associated through C−H···O bonds are more free to rotate around the bond.

4. Conclusions

The structural properties of liquid formic acid under normal conditions were investigated through both ab initio and classical molecular dynamics simulations, to highlight the determinant role of H-bonding in the mutual arrangement of molecular pairs and in the formation of a network of connected molecules. Our simulations show that the liquid is basically organized in two hierarchic structures. The lower level structure is characterized by small O−H···O bonded clusters, whose size does not exceed 20−30 molecules. At the second level, the clusters are held together by weak C−H···O bonds forming an extended network. Furthermore small H-bonded domains, in which about 20% of molecules participate, are embedded into the network without forming C−H···O bonds. Short H-bonded chains with no relevant branching dominate the H-bond network structure. In fact about 80% of the molecules is bound to one or two other neighbors by O−H···O bonds. This percentage drops to 65% when C−H···O bonds are also considered. Moreover, an analysis focused on the determination of cyclic H-bonded dimers has revealed that the ab initio simulation predicts a significant concentration (23% of molecules) of cyclic dimer configurations, typical of the gas phase. However several molecules (20%), belonging to cyclic dimers, are further H-bonded to other

molecules thus contributing to longer H-bonded chains. The resulting picture is that of an H-bond network of size comparable to the simulation box. To establish whether an infinite H-bond network exists, a much larger simulation box would be required. In contrast to other theoretical works,[11] the parallel pair molecular arrangement has been found to be relevant. This preferential orientation is shown to be basically due to the O−H···O bonds and in particular to cyclic H-bond dimer configurations. The C−H···O bonding instead is not found to favor a particular pair arrangement, the molecules being almost free to rotate around the H-bond.

Acknowledgment. We are grateful to Pál Jedlovszky for providing data from a reverse Monte Carlo simulation of liquid formic acid.[10] We also thank Gianni Cardini and Piero Procacci for stimulating discussions and Marco Pagliai for helpful insights on the ab initio simulation. This work was supported by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca, and by the European Union (grant no. RII3-CT-2003-506350).

References and Notes

(1) Almenningen, A.; Bastiansen, O.; Motzfeldt, T. *Acta Chem. Scand.* **1969**, *23*, 2848.
(2) Almenningen, A.; Bastiansen, O.; Motzfeldt, T. *Acta Chem. Scand.* **1970**, *24*, 747.
(3) Kwei, G.; Curl, R. *J. Chem. Phys.* **1960**, *32*, 1592.
(4) Nahringbauer, I. *Acta Crystallogr. B* **1978**, *34*, 315.
(5) Albinati, A.; Rouse, K. D.; Thomas, M. W. *Acta Crystallogr. B* **1978**, *34*, 2188.
(6) Nasr, S.; Bellissent-Funel, M.-C.; Cortès, R. *J. Chem. Phys.* **1999**, *110*, 10945.
(7) Bellamy, L. J.; Lake, R. F.; Pace, R. *J. Spectrochim. Acta* **1963**, *19*, 443.
(8) Waldstein, P.; Blatz, L. A. *J. Phys. Chem.* **1967**, *71*, 2271.
(9) Gorbunova, T.; Shilov, V. V.; Batalin, G. I. *Russ. J. Struct. Chem.* **1973**, *14*, 388.
(10) Jedlovszky, P.; Bakó, I.; Pálinkás, G.; Dore, J. C. *Mol. Phys.* **1995**, *86*, 87.
(11) Jedlovszky, P.; Turi, L. *J. Phys. Chem. B*, **1997**, *101*, 5429.
(12) Allan, R.; Clark, S. J. *Phys. Rev. Lett.* **1999**, *82*, 3464.
(13) Tomlinson, G. E.; Curnutte, B.; Hathaway, C. E. *J. Mol. Spectrosc.* **1970**, *36*, 26.
(14) Bartholomew, R. J.; Irsh, D. E. *J. Raman Spectrosc.* **1999**, *30*, 325.
(15) Hippler, M. *Phys. Chem. Chem. Phys.* **2002**, *4*, 1457.
(16) Pajak, Z.; Szczeœniak, E. *Chem. Phys. Lett.* **1977**, *49*, 269.
(17) Geisenfelder, H.; Zimmerman, H. *Ber. Bunsen-Ges. Phys. Chem.* **1963**, *67*, 480.
(18) Gulivets, N. I.; Lutskii, A. E.; Radchenko, I. V. *Russ. J. Struct. Chem.* **1965**, *6*, 20.
(19) Bakó, I.; Schubert, G.; Megyes, T.; Pálinkás, G.; Swan, G. I.; Dore, J.; Bellisent-Funel, M.-C. *Chem. Phys.* **2004**, *306*, 241.
(20) Constant, E.; Lebrun, A. *J. Chim. Phys.* **1964**, *61*, 163.
(21) Jedlovszky, P.; Turi, L. *J. Phys. Chem. A* **1997**, *101*, 2662.
(22) Mináry, P.; Jedlovszky, P.; Mezei, M.; Turi, L. *J. Phys. Chem. B* **2000**, *104*, 8287.
(23) Hermida Ramón, J. M.; Ríos, M. A. *Chem. Phys.* **1999**, *250*, 155.
(24) Chelli, R.; Cardini, G.; Procacci, P.; Righini, R.; Califano, S. *J. Chem. Phys.* **2002**, *116*, 6205.
(25) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.
(26) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(27) Hocking, W. H. *Z. Naturforsch A* **1976**, *31*, 1113.
(28) Riddick, J. A.; Bunger, W. B.; Sakano, T. K. *Organic Solvents*; Wiley: New York, 1986; vol. II.
(29) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993.
(30) Kleinman, L.; Bylander, M. *Phys. Rev. Lett.* **1982**, *48*, 1425.
(31) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(32) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(33) Hutter, J.; Alavi, A.; Deutch, T.; Bernasconi, M.; Goedecker, S.; Marx, D.; Tuckerman, M.; Parrinello, M. *CPMD* (MPI für Festkörperforschung and IBM Zurich Research Laboratory, Stuttgart, 1995−1999).
(34) Wang, B.; Hinton, J. F.; Pulay, P. *J. Phys. Chem. A* **2003**, *107*, 4683.
(35) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
(36) Kristyán, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175.
(37) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(38) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.5; Gaussian, Inc.: Pittsburgh, PA, 1998.
(39) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24.
(40) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
(41) Hoover, W. G. *Phys. Rev. A* **1986**, *34*, 2499.
(42) Marchi, M.; Procacci, P. *J. Chem. Phys.* **1998**, *109*, 5194.
(43) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
(44) Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990.
(45) Procacci, P.; Darden, T. A.; Paci, E.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848.
(46) Lide, D. R.; Frederiksen, H. P. R. *CRC Handbook of Chemistry and Physics*; CRC Press: Boca Raton, 1994−95.
(47) Kratochwill, A.; Hertz, H. G. *J. Chim. Phys.* **1977**, *74*, 814.
(48) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.
(49) Bakó, I.; Jedlovszky, P.; Pálinkás, G.; Dore, J. C. *Hydrogen Bond Networks*; Kluwer Academic: Dordrecht, 1994; vol. 435 (NATO ASI Series C), pp 119−127.
(50) Sindzingre, P.; Klein, M. L. *J. Chem. Phys.* **1992**, *96*, 4681.
(51) Root, L. J.; Berne, B. J. *J. Chem. Phys.* **1997**, *107*, 4350.
(52) Ferrario, M.; Haughney, M.; McDonald, I. R.; Klein, M. L. *J. Chem. Phys.* **1990**, *93*, 5156.
(53) Turi, L. *J. Phys. Chem.* **1996**, *100*, 11285.
(54) Bratoz, S.; Hadzi, D.; Sheppard, N. *Spectrochim. Acta* **1956**, *8*, 249.
(55) Shubin, A. A. *Izv. Akad. Nauk, Ser. Fiz.* **1950**, *14*, 442.
(56) Bertagnolli, H. *Chem. Phys. Lett.* **1982**, *93*, 287.
(57) Blumenfeld, M.; Fast, H. *Spectrochim. Acta, Part A* **1968**, *24*, 1449.
(58) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes*; Cambridge University Press: Cambridge, 1986.
(59) Brinkmann, N. R.; Tschumper, G. S.; Yan, G.; Schaefer, H. F., III. *J. Phys. Chem. A* **2003**, *107*, 10208.
(60) Kutzelnigg, W.; Klopper, W. *J. Chem. Phys.* **1991**, *94*, 1985.
(61) Gervasio, F. L.; Chelli, R.; Marchi, M.; Procacci, P.; Schettino, V. *J. Phys. Chem. B* **2001**, *105*, 7835.
(62) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. *J. Am. Chem. Soc.* **2002**, *124*, 6133.