Ab initio calculation of hydrogen bonds in liquids: A sequential Monte Carlo quantum mechanics study of pyridine in water

Thaciana Malaspina
Instituto de Física, Universidade de São Paulo, CP 66318, 05315-970, São Paulo, SP, Brazil

Kaline Coutinho
Universidade de Mogi das Cruzes, CP 411, 08701-970, Mogi das Cruzes, SP, Brazil

Sylvio Canuto[a)]
Instituto de Física, Universidade de São Paulo, CP 66318, 05315-970, São Paulo, SP, Brazil

(Received 12 February 2002; accepted 23 April 2002)

A systematic procedure based on the sequential Monte Carlo quantum mechanics (S-MC/QM) methodology has been used to obtain hydrogen bond strength and structures in liquids. The system considered is pyridine in water. The structures are generated by *NVT* Monte Carlo simulation, of one pyridine molecule and 400 water molecules. The hydrogen bonds are obtained using a geometric and energetic procedure. Detailed analysis shows that 62% of the configurations have one hydrogen bond. In the average, pyridine in liquid water makes 1.1 hydrogen bonds. The sampling of the structures for the quantum mechanical calculations is made using the interval of statistical correlation obtained by the autocorrelation function of the energy. A detailed statistical analysis is presented and converged results are obtained. The QM calculations are performed at the *ab initio* MP2/6-31+G(d) level and the results are compared with the optimized 1:1 cluster. Our results using QM calculations on 155 structures making one hydrogen bond gives an average binding energy of 3.7 kcal/mol, after correcting for basis set superposition error, indicating that in the liquid the binding energy is about 2/3 of the corresponding binding in the optimized cluster. © *2002 American Institute of Physics.* [DOI: 10.1063/1.1485963]

I. INTRODUCTION

Hydrogen bond continues to be a topic of increased interest in physics, chemistry, and biology.[1–6] It is fundamental to understand a series of biological phenomena that are known to occur only in water, including proteins, DNA and enzymatic reactions. Recent progress in cluster physical chemistry[5,7,8] with laser vaporization and jet-cooling techniques, have produced a rich variety of hydrogen-bonded complexes. These lead to an increasing interest in the theoretical and experimental investigation of both structure and spectroscopy of hydrogen-bonded systems.[9–28] However, the situation found in a cluster is, of course, not the same as that in a liquid and understanding the nature of hydrogen bonds in solution is crucial for several solvation and biomolecular processes. It has been recognized that extrapolating hydrogen bond energies from gas phase to aqueous is particularly risky.[6] Recent efforts have considered the solvent effect on the hydrogen bonds.[19,28] These studies have been concerned with the influence of the solvent environment.[19] Thus there are theoretical studies of the solvent (S) effects on a particular hydrogen-bonded system $A\cdots B$. The $(A\cdots B)+S$ is studied using continuum theories. Lischka and co-workers[19] have studied hydrogen-bonded interactions in the acetic acid dimer and in complexes formed by acetic acid in different solvents using the polarized continuum model[29–37] for the description of the solvent. In this investigation we are interested in the situation where the hydrogen bond is formed directly with the solvent, $A\cdots S$. This specific interaction is analyzed with the solvent as a dynamic, statistical, liquid at a certain temperature. The difficulty in the theoretical study of hydrogen bonds in liquids is associated to the fact that in a liquid there is not one, or even a few, but a very large number of different structures at a certain temperature. Although there has been some effort to estimate theoretically the hydrogen bond strength in liquids,[20] using results from systematic cluster calculations, it still lacks the statistical nature that is intrinsic in a liquid system. Monte Carlo or molecular dynamics computer simulation can give the essential aspects of the structure of a liquid and this type of treatment has indeed been successfully used in obtaining the solvent effects in the absorption UV/visible spectrum.[38–46] In our case a sequential treatment of the QM/MM (quantum mechanics/molecular mechanics) type is used. The structure of the liquid is generated by Monte Carlo computer simulation and quantum mechanical calculations are performed next in these super-molecular structures. In this sequential (S-MC/QM) treatment all molecules are treated by quantum mechanics. In this present paper this methodology is extended to obtain hydrogen bond strength in liquids. To the best of our knowledge, this is the first *ab initio* theoretical calculation of the specific interaction hydrogen bond between the solute and the solvent that explicitly considers the statistical average that characterizes the liquid. The calculated final result represents an average of structures obtained for the liquid in the thermodynamic equilibrium. The system considered here is the pyridine–water system. The selection of pyridine is mo-

[a)]Author to whom correspondence should be addressed. Electronic mail: canuto@if.usp.br

tivated by the strong interest involved in six-membered nitrogenated aromatic rings as the building block of proteins and nucleotides. Theoretical studies of the 1:1 complex of pyridine-water have been performed previously[21–24] aiming at understanding the cluster situation. As we shall see, in the case of pyridine in liquid water more than one hydrogen bond can be formed and this type of statistic is of interest. Monte Carlo simulation is performed for one pyridine molecule in water at room temperature. Then hydrogen-bonded structures in the liquid are identified and separated. These structures are submitted to single-point *ab initio* second-order perturbation theory (MP2) calculations and the binding energies are then obtained and averaged. Another aspect that has to be considered in the theoretical calculations is the effect of basis set superposition error (BSSE).[47] All results presented here are fully corrected for BSSE using the counterpoise correction.[47,48] Thus several quantum mechanical calculations are performed on the structures generated from the Monte Carlo simulation and all results are corrected for BSSE. The binding energy is thus obtained using configurational averages of several hydrogen bonded structures obtained from Monte Carlo simulation at $T=25\,^{\circ}\mathrm{C}$. The number of structures used is obtained sampling the MC structures after analysis of the interval of statistical correlation.[38,39]

Estimating the difference between the hydrogen bond in gas and in liquid is important also in the context of classification of the amount of covalence involved.[6] In a liquid it is expected that the hydrogen bond is weaker compared to the same binding in a cluster.[3,6,19,49] This paper describes an *ab initio* attempt to estimate this difference in hydrogen bond strength.

II. MONTE CARLO SIMULATION

Monte Carlo (MC) statistical mechanics simulations are carried out employing standard procedures,[50] including the Metropolis sampling technique[51] and periodic boundary conditions using the minimum image method in a cubic box. The simulations are performed in the canonical (NVT) ensemble. The system consists of one pyridine molecule plus 400 water molecules. The volume of the cubic box is determined by the experimental[52] density of water, which at $T=298.15\,\mathrm{K}$ is $0.9966\ \mathrm{g/cm^3}$. The intermolecular interactions are described by the standard Lennard-Jones plus Coulomb potential with three parameters for each atom i (ϵ_i, σ_i, and q_i). The atomic parameters are combined by geometric average to generate the pair parameters $\epsilon_{ij}=(\epsilon_i\epsilon_j)^{1/2}$ and $\sigma_{ij}=(\sigma_i\sigma_j)^{1/2}$. For the water molecules we use the SPC potential developed by van Gunsteren *et al.*[53] For pyridine, we use two models of OPLS potential, the 6-site[54] and 11-site.[55] Most of the quantum mechanical calculations are made on the structures obtained by the MC simulation using the 11-site model. However, comparison with the 6-site model indicates the possible dependence of the results on the particular choice of the classical potential. The set of intermolecular potential parameters employed in the simulations are shown in Table I. The intermolecular interactions are spherically truncated within a center of mass separation smaller than the cutoff radius, r_C, of 11.5 Å. Long range corrections were calculated beyond this cutoff distance.[50] The Lennard-Jones potential contribution is estimated assuming an uniform distribution $G(r)\approx 1$ after the cutoff radius and the electrostatic potential contribution is estimated with the reaction field method of the dipolar interaction. In the simulation the molecules are kept with rigid geometries. During the classical simulation the water molecules are kept in their C_{2v} structure with $r_{\mathrm{OH}}=1.000$ Å and $\theta(\hat{\mathrm{HOH}})=109.47°$. The quantum mechanical calculations, however, have a relaxed geometry for the water. The pyridine molecule is also held rigid in its structure, optimized at the cluster (1:1) with MP2/6-31 $+\mathrm{G}(d)$ level of theory. These geometries are discussed in the next section when we present the quantum mechanics methodology.

TABLE I. Potential parameters used in the Monte Carlo simulations (q in elementary charge unit, ϵ in kcal/mol and σ in Å).

System	q	ϵ	σ
Pyridine 6-site[a]			
N_1	−0.49	0.170	3.250
C_2H_7, C_6H_{11}	0.23	0.110	3.750
C_3H_8, C_5H_{10}	−0.03	0.110	3.750
C_4H_9	0.09	0.110	3.750
Pyridine 11-site[b]			
N_1	−0.678	0.170	3.250
C_2, C_6	0.473	0.070	3.550
C_3, C_5	−0.447	0.070	3.550
C_4	0.227	0.070	3.550
H_7, H_{11}	0.012	0.030	2.420
H_8, H_{10}	0.155	0.030	2.420
H_9	0.065	0.030	2.420
Water SPC[c]			
O	−0.820	0.155	3.165
H	0.410	0.000	0.000

[a]Reference 54.
[b]Reference 55.
[c]Reference 53.

The initial configurations are generated randomly, considering the position and orientation of each molecule. One MC step is performed after one water molecule randomly attempts to translate in the Cartesian directions and also attempts to rotate around a randomly chosen axis. The maximum allowed displacement of the molecules is self-adjusted after 50 configurations to give an acceptance ratio of new configurations around 50%. The maximum rotation angle was fixed during the simulation in $\delta\theta=\pm 15°$. The simulations consisted of a thermalization phase of 2.0×10^6 MC steps, followed by an averaging stage of 80.0×10^6 MC steps, where the thermodynamical properties are evaluated and the configurations are generated.

As quantum mechanical calculations will be performed on the configurations generated by the MC simulation it is important to optimize the statistics. Successive configurations, that are statistically highly correlated, will not give important additional information. Therefore we calculate the interval of statistical correlation using the autocorrelation function of the energy.[56–58] Essentially, it gives the interval of MC steps that statistically uncorrelated configuration can be sampled. For the simulations presented here, we obtain that configurations separated by 320×10^3 MC steps are sta-

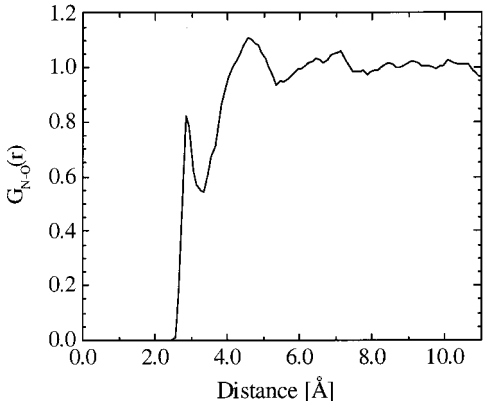

FIG. 1. The calculated radial distribution function between the nitrogen atom of the pyridine and oxygen atom of the water molecules, $G_{\mathrm{N-O}}(r)$.

tistically uncorrelated. Then, the total of 80×10^6 successive configurations generated in the simulation can be drastically reduced to 250 uncorrelated configurations whitout loss of statistical information.[41,57] Thus, after each simulation 250 uncorrelated configurations were sampled to be used in further quantum mechanical supermolecular calculations. A detailed discussion on this procedure of sampling configurations is given below in Sec. V.

All simulation are performed with the DICE[59] Monte Carlo statistical mechanics program. DICE is a general program for MC simulation with a graphical interface that calculates thermodynamic properties and generates structures for using in most conventional quantum chemistry program.

III. STATISTICAL ANALYSIS

A. Hydrogen bonds from Monte Carlo simulation

An important point in this study is the identification of the hydrogen bonds obtained in the Monte Carlo simulation. We first analyze the radial distribution function because it is the conventional and well known procedure to give the coordination number. Figure 1 shows the radial distribution function, $G(r)$, between the nitrogen of pyridine and the oxygen of water, $G_{\mathrm{N-O}}(r)$ for the 11-site potential model used to simulate the pyridine. The hydrogen bonds are obtained from the analysis of this $G_{\mathrm{N-O}}(r)$, that has a well-defined first peak, indicating a hydrogen-bond structure. The first peak in the $G_{\mathrm{N-O}}(r)$ distribution function starts at 2.45 Å and ends at 3.35 Å, with a maximum at 2.94 Å. The spherical integration of this first peak in the $G_{\mathrm{N-O}}(r)$ over the corresponding interval, gives 1.76 water molecules as nearest neighbors. The uncertainty associated with this procedure is that it can not be assured that all nearest-neighbor structures involved with a distance N–O smaller than the minimum of $G_{\mathrm{N-O}}(r)$ (3.25 Å in this case) are indeed associated with hydrogen bonds. Associating this in combination with the N–H distribution gives better results but structures that cannot be associated to hydrogen bond still persist. A more efficient and correct way to extract the hydrogen-bonded structures can be used. Stilinger and Rahman[60–62] and Mezei and Beveridge[63] have discussed the directional and energetic aspects of hydrogen bonds and its usefulness in identifying hydrogen bonds in computer simulation of liquids.

TABLE II. Statistics of the hydrogen bonds formed between pyridine (6- and 11-site) and water. Uncertainties are the standard deviation.

	6-site	11-site
	Percentage of configurations	
0 hydrogen bonds	24%	17%
1 hydrogen bonds	63%	62%
2 hydrogen bonds	13%	20%
3 hydrogen bonds	0%	1%
$\langle R_{\mathrm{N-O}}\rangle$ (Å)	3.04±0.21	2.94±0.16
$\langle\theta(\mathrm{N\hat{O}H})\rangle$ (degree)	17±7	14±7

Hydrogen bonds are better obtained using the geometric and energetic criteria.[38,60–64] We consider a hydrogen bond formation when the distance $R_{\mathrm{N-O}}\leqslant3.5$ Å the angle $\theta(\mathrm{N\hat{O}H})\leqslant30°$ and the binding energy is positive. In doing so, in the 250 MC configurations we find 223 and 269 hydrogen bonds formed in the nitrogen atom of pyridine in the 6- and 11-site potentials. This gives an average of 0.9 and 1.1 hydrogen bonds, for the 6-site and 11-site potentials, respectively. Note that these are averages for the liquid. Pyridine can form up to three hydrogen bonds with the surrounding liquid water molecules and Table II gives the complete statistics obtained for the hydrogen bonds formed. Note the detailed information that can be obtained with this procedure and it cannot be done in a simple fashion using only the $G(r)$ distribution function. There is a clear predominance of the structures with one hydrogen bond and this result is not very much dependent on the potential. In the 6-site model we find that 63% of the configurations form one hydrogen bond. This is to be compared with the result of 62% for the 11-site potential. We find that, for instance in the 11-site potential, in 17% of the configurations the pyridine does not form any hydrogen bonds, in 62% it forms one, and in 20% forms two and in only 1% it forms three hydrogen bonds. The average quantum mechanical values computed for the hydrogen-bonding energies in the liquid are obtained using these structures. There is a total of 155 configurations making one hydrogen bond. All these 155 structures composed of one pyridine and one water will be submitted to quantum mechanical calculations of the binding energies.

As the appropriate Boltzmann weights are included in the Metropolis Monte Carlo sampling technique[50] the average value of the binding energy, or any other property calculated from the MC data, is given as a simple average over a chain of size L of energy values:

$$\langle E\rangle=\frac{1}{L}\sum_{i}^{L}E_i. \tag{1}$$

It is known that independent, or *uncorrelated*, values of E_i generate a normal distribution with a standard deviation (s) given by

$$s=\sqrt{\frac{L}{(L-1)}(\langle E^2\rangle-\langle E\rangle^2)} \tag{2}$$

and a statistical error of the average (σ) given simply by

$$\sigma = \frac{s}{\sqrt{(L-1)}}. \tag{3}$$

Note that for large L the standard deviation converges to a constant value and the statistical error tends to zero. Thus, the final value represented by $\langle E \rangle \pm s$ shows the two informations that describe the normal distribution and comprises 68% of the data.

B. Sampling configurations

The question of sampling configurations from statistical simulations has been recognized to be a very important issue and it is crucial for the efficiency of QM/MM methods. Instead of performing a quantum mechanical calculation on every configuration generated by the MC simulation, we use the interval of statistical correlation and the statistical inefficiency, to select the configuration that gives relevant statistical information.[39,43,56–58] We have shown numerically that performing average over thousands of successive configurations generated in the MC simulation, gives the same result as averaging over only a few statistically uncorrelated configurations.[43,57] This is a very efficient way to sample configurations. It is statistically sound and, as we shall see below it gives statistically converged results. In doing so, the subsequent quantum mechanical calculations are performed only on some uncorrelated structures. This is one of the advantages of the sequential procedure of the S-MC/QM, in that all the important MC statistical informations are available before running into the QM calculations. As in previous works[38–43,58] we calculate the autocorrelation function of the energy, $C(n)$, to obtain the interval of statistical correlation, using the definition

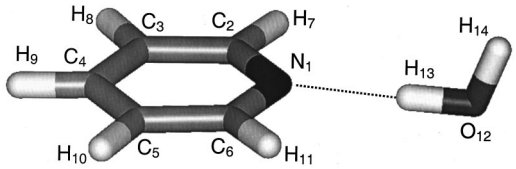

FIG. 2. The structure of pyridine–water complex. Atomic numbering is used to define the geometrical parameters.

$$C(n) = \frac{\langle \delta E_i E_{i+n} \rangle}{\langle \delta E^2 \rangle} = \frac{\Sigma_i (E_i - \langle E \rangle)(E_{i+n} - \langle E \rangle)}{\Sigma_i (E_i - \langle E \rangle)^2}, \tag{6}$$

where E_i is the energy of a configuration i and E_{i+n} is the energy of the configuration generated n MC step later. For Markovian processes, it is known that $C(n)$ follows an exponential decay[65] $C(n) = \exp(-n/\tau)$ and represents the statistical correlation between configurations separated by n MC steps. Thus, $C(n) = 1$ means that configurations separated by n MC steps are 100% statistically correlated and do not contribute with new statistical information to the average. Analyzing the exponential decay is easy to see that only with an infinite separation the configurations will be statistically uncorrelated, $C(n = \infty) = 0$. However, in practice the configurations are considered statistically uncorrelated for an interval $n \approx 2\tau$ where $C(n) \cong 0.13$, i.e., less that 13% of statistical correlation. In the simulations of molecular liquids performed by us, the $C(n)$ was best described by the sum of two exponential functions, $C(n) = c_1 \exp(-n/\tau_1)$

TABLE III. The optimized geometry of the pyridine–water cluster (1:1) in comparison with the isolated moieties using the MP2/6-31+G(d) level. Also shown are the experimental data for isolated pyridine (Ref. 70) and water (Ref. 71). Atomic indices are defined in Fig. 2.

Distance (Å)	Isolated	Cluster (1:1)	Experiment
N_1C_2 and N_1C_6	1.347	1.347	1.338
C_2C_3 and C_5C_6	1.398	1.396	1.394
C_4C_3 and C_4C_5	1.396	1.396	1.392
C_2H_7 and C_6H_{11}	1.088	1.088	1.086
C_3H_8 and C_5H_{10}	1.087	1.087	1.082
C_4H_9	1.088	1.088	1.081
N_1H_{13}		1.964	
N_1O_{12}		2.947	
$O_{12}H_{13}$	0.971	0.984	0.957
$O_{12}H_{14}$	0.971	0.970	0.957
Angle (degree)			
$C_2N_1C_6$	116.8	117.6	116.9
$N_1C_2C_3$ and $N_1C_6C_5$	123.7	123.2	123.8
$C_2C_3C_4$ and $C_6C_5C_4$	118.6	118.7	118.5
$C_3C_4C_5$	118.4	118.6	118.4
$N_1O_{12}H_{13}$		0.7	
$H_{13}O_{12}H_{14}$	105.5	105.2	104.5
Dihedral (degree)			
$H_{13}N_1C_2C_3$		171.6	
$H_{13}N_1C_2H_7$		−8.3	
$H_{14}O_{12}N_1C_2$		96.2	
$H_{14}O_{12}N_1C_6$		−92.8	

$+c_2\exp(-n/\tau_2)$, where $\tau_1 \gg \tau_2$. In this case, τ is calculated by integrating $C(n)$ from zero to infinite, then $\tau = c_1\tau_1 + c_2\tau_2 \cong c_1\tau_1$. The same behavior of the autocorrelation function was found also by other authors[66–68] in simulations of spin models in a lattice.

For pyridine in water we obtain here that configurations separated by 320×10^3 MC steps are statistically uncorrelated. Using configurations separated by less than this is a waste because it includes configurations that do not contribute to the average. Further, if the simulation is not long enough it will give statistically unconverged results in spite of the large computational effort. Therefore, we select one configuration in each 320×10^3 MC steps and use them to perform QM calculations. This assures that the structures used in the quantum mechanical calculations are statistically relevant and converged values are obtained, as it will be demonstrated below. As the total number of MC steps in the simulation was 80×10^6, the averages, including the quantum mechanical calculations, are then taken over 250 uncorrelated configurations ($250 = 80\times10^6/320\times10^3$).

IV. QUANTUM MECHANICAL CALCULATIONS

A. The optimized 1:1 cluster

All quantum mechanical calculations are performed using GAUSSIAN 98[69] at the *ab initio* second-order perturbation theory, MP2, level thus including electron correlation effects in all results. As a very large number of calculations have to be performed (nearly 1000) to keep the computational effort within a reasonable limit it is necessary to make some analysis of the influence of the basis set. The basis set employed here is the double-zeta plus diffuse and polarization, 6-31 +G(d). In a recent investigation Dkhissi, Adamowicz, and Maes[24] presented a comparative study of several different theoretical methods for the 1:1 pyridine–water complex. Our result for the binding energy of the optimized 1:1 cluster with MP2/6-31+G(*d*) is 6.24 kcal/mol after correcting for BSSE. This is in very good agreement with their equivalent result of 6.29 kcal/mol.[24] Frequency calculation is also performed to assure that the geometry obtained is indeed a true minimum of the energy. The optimized cluster is shown in Fig. 2. This is the same structure as obtained previously,[21,24] with the hydrogen atom of water binding to the nitrogen atom of pyridine. The other hydrogen atom of water points perpendicular to the aromatic plane of pyridine. In the experimental studies of the hydrogen bonded clusters formed between diazines and water, Caminati and co-workers[17,18] obtained the water located parallel to the aromatic ring. Here, in the pyridine case, this possible energy-minimum structure was not found. Similar to previous studies[21,24] the plane formed by water and pyridine are nearly perpendicular. The dihedral angle $H_{14}O_{12}N_1C_2$ (see Fig. 2) is calculated as 96°. In Table III, the optimized geometry of the 1:1 cluster is shown in comparison with the isolated moieties and the experimental data of the monomers.[70,71] The hydrogen bond distance N_1-H_{13} and N_1-O_{12} are calculated as 1.964 and 2.947 Å, respectively. These are typical distances of a medium strength hydrogen bond. In the MC simulation the maximum of G_{N-O} is located at 2.94 Å (Fig. 1). Comparing

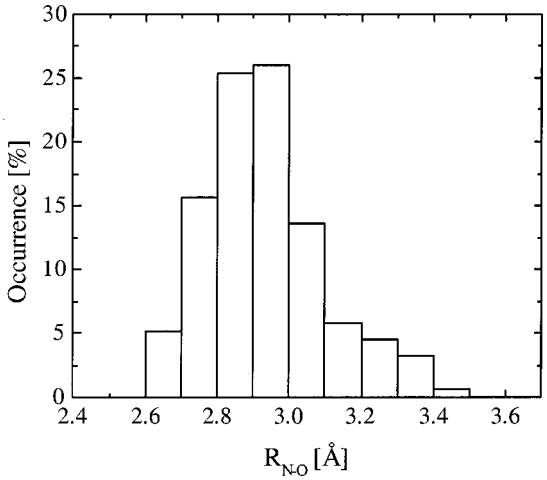

FIG. 3. Histogram of hydrogen bond distances obtained from the MC simulation.

the cluster situation with the liquid it can be noted that there are no marked differences in the average N–O distance. With the 11-site potential the average value is 2.94 Å (Table II), compared to the optimized cluster value of 2.95 Å. The 6-site potential gives a slightly longer distance, which is possibly a consequence of the smaller classical charge on the N atom. However, note that the standard deviation is also larger for the 6-site potential. In any case the N–O distance calculated for the 1:1 cluster is within the statistical result for the liquid 2.94 ± 0.16 or 3.04 ± 0.21 Å. To give a clear picture of the spread Fig. 3 shows the histogram of the N–O distance. The histogram is asymmetric and reflects the asymmetry of the interatomic potential. The potential is more repulsive for lower distances than it is for larger distances. Figure 4 shows some of the structures derived from the simulation [Figs. 4(b), 4(c), and 4(d)], in comparison with the optimized 1:1 cluster [Fig. 4(a)].

The largest geometrical relaxation upon complexation occurs in the water molecule. As it can be seen in Table III, the water O–H distance involved in the hydrogen bond OH···N is increased by 0.013 Å. Again, this stretch, is in good agreement with the previous *ab initio* MP2 results.[21,24] In the pyridine moiety one can note a slight increase in the $C_2N_1C_6$ angle. This lengthening of the $O_{12}H_{13}$ distance is responsible for the redshift of the OH stretching vibrations. The hydrogen bond also leads to a blueshift of the bending

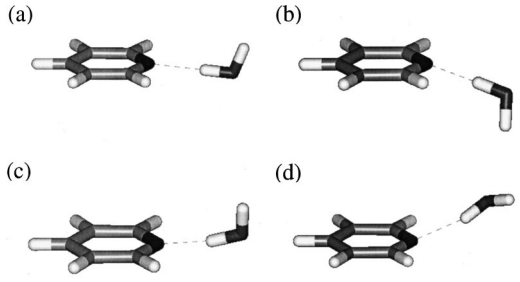

FIG. 4. Illustration of some hydrogen bonded structures generated in the MC simulation and comparison with the optimized cluster shown in (a).

TABLE IV. The rotational constants and the dipole moment of the optimized geometry of the pyridine-water cluster (1:1) in comparison with the isolated moieties using MP2/6-31+G(d) level. Calculated average dipole moment for the liquid structures is 4.72 D.

Property	Isolated Present	Isolated Ref. 24	Expt. (Ref. 70)	Cluster (1:1) Present
Rotational constants (MHz)	6005.9	6015.5	6039.2	5902.2
	5772.3	5777.7	5804.9	1405.9
	2943.4	2947.1	2959.2	1141.3
Dipole (Debye)	2.51	2.46	2.15	5.07

vibrational mode. This is very well described here. The symmetric and asymmetric stretching modes are calculated to be redshifted by 185 cm^{-1} and 54 cm^{-1}, whereas the bending mode is calculated to blue shift by 47 cm^{-1}. In Table IV we compare the calculated rotational constants and dipole moment with experiment and previous theoretical results. The MP2/6-31+G(d) dipole moment calculated for the 1:1 optimized cluster is 5.07 D, compared to the average dipole moment of 4.72 D for the pyridine–water using the structures of the liquid.

B. 1:1 cluster in liquid

The major interest of this paper is the calculation of the binding energy of pyridine in liquid water and a comparison with the result obtained for the optimized 1:1 cluster. All binding energies are obtained using counterpoise correction to BSSE, i.e.,

$$\Delta E_{A-B}^{\mathrm{CP}} = E_{A-B} - E_{A(B)} - E_{B(A)}, \tag{4}$$

$$\Delta E_{A-B} = E_{A-B} - E_A - E_B, \tag{5}$$

where, as usual, E_{A-B} is the energy of the complex, $E_{A(B)}$ is the energy of the monomer A obtained with the entire basis set, including the basis set of monomer B, and $E_{B(A)}$ is the equivalent for the monomer B. E_A and E_B are the energies of the monomers A and B. Thus the difference between ΔE^{CP} and ΔE gives the basis set superposition error. It is interesting to compare this error in the cluster and in the liquid. As discussed before 250 configurations were extracted from the Monte Carlo simulation. In these we separate the 155 configurations (62%) with one hydrogen bond using the structure obtained by the MC simulation. Thus we calculate the binding energies of one hydrogen bond, with and without the counterpoise corrections, using Eqs. (4) and (5) 155 times, leading to a total of 775 QM calculations of the MP2/6-31 +G(d) level.

The calculated binding energies are summarized in Table V. There are several things worth noting on these results. First, it can be noted that the counterpoise correction to BSSE is similar in the cluster and in the liquid, both amounting to 2.1 kcal/mol. There is no reason why one should expect *a priori* that these numbers are the same, if we recall that in the liquid the pyridine-water shows a large variety of conformations with several N–O distances and relative orientations. However, on the average, for this case, the counterpoise correction is the same as in the 1:1 cluster. The magnitude of the binding energy is expected to decrease in the liquid compared to the optimized cluster. After the counterpoise correction to the BSSE the cluster binding energy is calculated as 6.24 kcal/mol, in good agreement with a recent *ab initio* study.[24] In the liquid, we obtain the average binding energy using the structures generated and separated by the MC simulation as 3.70 kcal/mol, around 2.5 kcal/mol weaker than the same binding in the cluster. Now we can go slightly beyond and analyze the contribution of the structures with two hydrogen bonds. This relates to the well-known cooperative effect.[3,16,72–74] To have an indication of this hydrogen bond magnitude we have considered the cases of the configurations with two hydrogen bonds that correspond to 20% of the configurations generated in the MC simulation. A similar analysis (51 structures with two hydrogen bonds) shows that due to the second water the binding energy is 0.15 kcal/mol higher, in the average. We thus find here that the binding energy in the liquid is about 2/3 of the binding calculated in the optimized cluster.

V. CONVERGENCE OF THE RESULTS

As the calculated value obtained for the liquid case is derived from several quantum mechanical MP2 calculations using the structures of the simulation, it is important now to discuss the convergence of the final result. It is also convenient to show the statistical efficiency obtained with the au-

TABLE V. Calculated binding energy of hydrogen bond of pyridine and water obtained with MP2/6-31 +G(d) including counterpoise correction to BSSE. Results shown for the liquid are average values and the statistical errors are also shown. The standard deviation is given in parentheses.

Binding energy (kcal/mol)	Present results MP2/6-31+G(d)	Ref. 24 MP2/6-31++G(d,p)	Present results MP2/6-31+G(d)
	Cluster (1:1)	Cluster (1:1)	Liquid
ΔE	8.36	7.89	5.84±0.10 (1.20)
ΔE^{CP}	6.24	6.29	3.70±0.10 (1.24)

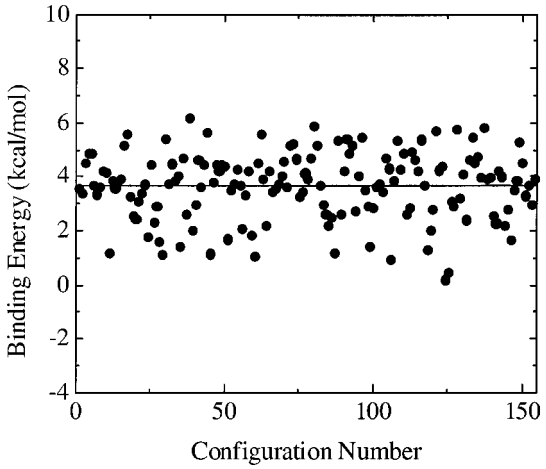

FIG. 5. Distribution of the individual quantum mechanical MP2 values of the hydrogen binding energies of pyridine and water obtained using the structures generated by the MC simulation.

tocorrelation function of the energy. Figure 5 shows the snapshots of the calculated values for all the 155 configurations used. This spread of values is a characteristic of the liquid and cannot be described by a single cluster. Figure 6, in complement, shows the histograms of these same calculated values. Note that 68% of the calculated energies are within the range of 2.46 kcal/mol and 4.94 kcal/mol, corresponding to the limits of $\langle E\rangle \pm s$; and 95% are within the limits of $\langle E\rangle \pm 2s$. At this stage it may be convenient to analyze the dependence of our calculated average value of the hydrogen binding energy on the set of L values used in the calculation. As discussed before the configurations separated from the MC simulation are statistically uncorrelated and therefore the average value should converge fast and systematically and should be independent of the particular choice of the set of MC configurations selected for the QM calculations. The statistical error of the average, however, depends on the total number of uncorrelated configurations (L) used to calculate the average; i.e., on the total size of the simulation. Figure 7 shows the calculated average binding energy for different L. The results clearly demonstrate that the average value has converged after around 80 QM binding

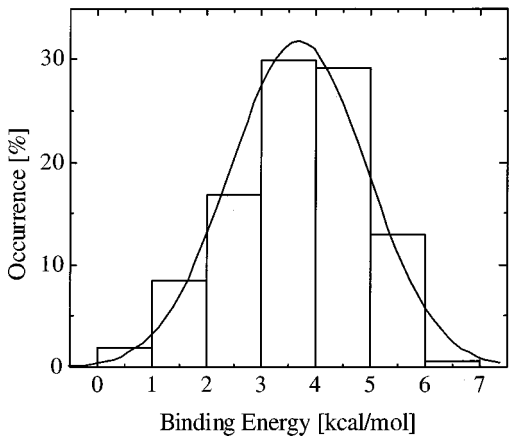

FIG. 6. Histogram of calculated quantum mechanical binding energies.

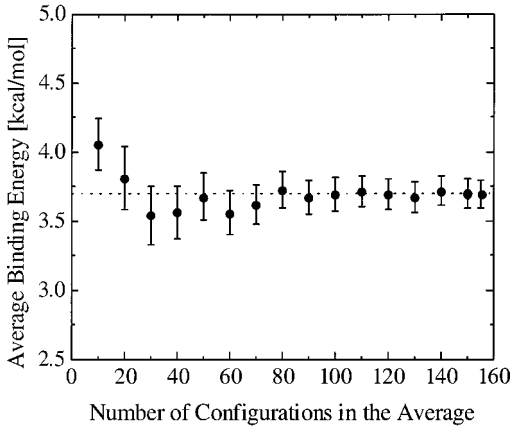

FIG. 7. Convergence of the calculated average binding energy.

energy calculations. This is again a demonstration of the efficiency of the use of the autocorrelation function in that only $L \approx 80$ instantaneous values give statistically converged results. Increasing further the number of calculations does not change the average binding energy but, of course, the statistical error σ decreases with increasing L, as it can be seen in Eq. (3); the standard deviation s, instead goes to a constant value. In complement, it can be seen that using less than 80 QM calculations leads to nonconverged results. That would correspond to a simulation with less than 40×10^6 MC steps. Thus the use of the interval of the statistical correlation (2τ) obtained using the autocorrelation function of the energy is a very effective way to assure statistically converged results with a relatively small number of calculations, because only statistically uncorrelated configurations are included. Our present results for the hydrogen bond between pyridine and water is a converged value incorporating the inherent statistical nature of the liquid. Our best result indicates that the binding energy for the pyridine in liquid water is 2/3 of the corresponding binding in the cluster. To our knowledge this is the first *ab initio* statistical analysis of the binding energy of a hydrogen bond in a liquid system.

VI. SUMMARY AND CONCLUSIONS

A very systematic procedure based on the sequential Monte Carlo quantum mechanics (S-MC/QM) methodology has been used to obtain hydrogen bond strength and structures in liquids. Using the structures generated by Monte Carlo simulation, QM calculations are performed at the *ab initio* MP2/6-31+G(d) level and the results are compared with the equivalent binding in the optimized 1:1 cluster. The hydrogen bond structures are obtained using a geometrical and energetic procedure that is a more general definition than the procedure using the coordination number generated in the radial distribution function. Detailed analysis shows that there are structures that make no hydrogen bonds and a few others that make two hydrogen bonds. 62% of the configurations have one hydrogen bond. In the average, pyridine in liquid water makes 1.1 hydrogen bonds. The sampling of the structures for the quantum mechanical calculations is made using the interval of statistical correlation obtained by the autocorrelation function of the energy. Converged results are

obtained using a little more than 80 structures. Our results using QM calculation on 155 structures making one hydrogen bond gives an average binding energy of 3.7 kcal/mol, after correcting for basis set superposition error, indicating that in the liquid the binding energy is about 2/3 of the corresponding binding in the optimized cluster. This suggests that the use of optimized cluster to mimic situations of the liquid in specific interaction between solute and solvent is not recommended. The methodology presented here can of course be used in estimating other situations and systems and the present application demonstrates the feasibility of this approach. The S-MC/QM methodology treats all molecules, solute and solvent, by quantum mechanics. In addition, running first the MC simulation gives important statistical information that is advantageously used in the subsequent QM calculations.

ACKNOWLEDGMENTS

This work has been partially supported by CNPq, CAPES, and FAPESP.

[1] G. C. Pimentel and A. L. McGellan, Annu. Rev. Phys. Chem. **22**, 347 (1971).
[2] G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures* (Springer-Verlag, Berlin, 1991).
[3] S. Scheiner, *Hydrogen Bonding: A Theoretical Perspective* (Oxford University Press, Oxford, 1997).
[4] B. M. Ladanyi and M. S. Skaf, Annu. Rev. Phys. Chem. **44**, 335 (1993).
[5] T. S. Zwier, Annu. Rev. Phys. Chem. **47**, 205 (1996).
[6] C. L. Perrin and J. B. Nielson, Annu. Rev. Phys. Chem. **48**, 511 (1997).
[7] *The Chemical Physics of Atomic and Molecular Clusters*, edited by G. Scoles (North-Holland, Amsterdam, 1990).
[8] *Atomic and Molecular Clusters*, edited by E. R. Bernstein (Elsevier, Amsterdam, 1990).
[9] A. D. Buckingham, P. W. Fowler, and J. M. Hutson, Chem. Rev. **88**, 963 (1988).
[10] I. V. Brovchenko and A. V. Oleinikova, J. Chem. Phys. **106**, 7756 (1997).
[11] G. T. Evans, J. Chem. Phys. **115**, 1440 (2001).
[12] M. M. Szczesniak, Z. Latajka, and S. Scheiner, J. Mol. Struct.: THEOCHEM **135**, 179 (1986).
[13] S. Scheiner, T. Kar, and Y. Gu, J. Biol. Chem. **276**, 9832 (2001).
[14] P. Hobza and Z. Havlas, Chem. Rev. **100**, 4253 (2000).
[15] R. Rivelino and S. Canuto, Chem. Phys. Lett. **322**, 207 (2000).
[16] R. Rivelino and S. Canuto, J. Phys. Chem. A **105**, 11260 (2001).
[17] W. Caminati, P. Moreschini, and P. G. Favero, J. Phys. Chem. A **102**, 8097 (1998).
[18] S. Melandri, M. E. Sanz, W. Caminati, P. G. Favero, and Z. Kisiel, J. Am. Chem. Soc. **120**, 11504 (1998).
[19] A. J. A. Aquino, D. Tunega, G. Haberhauer, M. H. Gerzabek, and H. Lischka, J. Phys. Chem. A **106**, 1862 (2002).
[20] D. M. Opadhyay, M. K. Shukla, and P. C. Mishra, Int. J. Quantum Chem. **81**, 90 (2001).
[21] M. A. Martoprawiro and G. B. Bacskay, Mol. Phys. **85**, 573 (1995).
[22] G. Maes, J. Smets, L. Adamowicz, W. McCarthy, M. K. Van Bael, L. Houben, and K. Schoone, J. Mol. Struct. **410**, 315 (1997).
[23] J. Smets, W. McCarthy, G. Maes, and L. Adamowicz, J. Mol. Struct. **476**, 27 (1999).
[24] A. Dkhissi, L. Adamowicz, and G. Maes, J. Phys. Chem. A **104**, 2112 (2000).
[25] G. Gilli and P. Gilli, J. Mol. Struct. **552**, 1 (2000).
[26] J. E. Del Bene, W. B. Person, and K. Szczepaniak, J. Phys. Chem. **99**, 10705 (1995).
[27] B. M. Ladanyi and M. S. Skaf, J. Mol. Struct.: THEOCHEM **335**, 181 (1995).
[28] S. Kawahara, T. Wada, S. Kawauchi, T. Uchimaru, and M. Sekine, J. Phys. Chem. A **103**, 8516 (1999).
[29] O. Tapia and O. Goscinski, Mol. Phys. **29**, 1653 (1975).
[30] J. L. Rivail and D. Rinaldi, Chem. Phys. **18**, 233 (1976).
[31] S. Miertus, E. Scrocco, and J. Tomasi, J. Chem. Phys. **55**, 117 (1981).
[32] M. M. Karelson and M. C. Zerner, J. Phys. Chem. **96**, 6949 (1992).
[33] A. Klamt and G. Schürman, J. Chem. Soc., Perkin Trans. 2 **2**, 799 (1993).
[34] J. Tomasi and M. Persico, Chem. Rev. **94**, 2027 (1994).
[35] C. J. Cramer and D. G. Truhlar, Chem. Rev. **99**, 2161 (1999).
[36] F. J. Luque, J. M. López, and M. Orozco, Theor. Chem. Acc. **103**, 343 (2000).
[37] J. Tomasi, Theor. Chem. Acc. **103**, 196 (2000).
[38] S. Canuto and K. Coutinho, Int. J. Quantum Chem. **77**, 192 (2000).
[39] K. Coutinho, S. Canuto, and M. C. Zerner, J. Chem. Phys. **112**, 9874 (2000).
[40] K. Coutinho and S. Canuto, J. Chem. Phys. **113**, 9132 (2000).
[41] K. J. de Almeida, K. Coutinho, W. B. de Almeida, W. R. Rocha, and S. Canuto, Phys. Chem. Chem. Phys. **3**, 1583 (2001).
[42] W. R. Rocha, K. J. de Almeida, K. Coutinho, and S. Canuto, Chem. Phys. Lett. **345**, 171 (2001).
[43] W. R. Rocha, K. Coutinho, W. B. de Almeida, and S. Canuto, Chem. Phys. Lett. **335**, 127 (2001).
[44] J. T. Blair, K. Krogh-Jespersen, and R. M. Levy, J. Am. Chem. Soc. **111**, 6948 (1989).
[45] J. Zeng, N. S. Hush, and J. R. Reimers, J. Chem. Phys. **99**, 1496 (1993).
[46] J. Gao, J. Am. Chem. Soc. **116**, 9324 (1994).
[47] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, Chem. Rev. **94**, 1873 (1994).
[48] S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).
[49] K. Coutinho, N. Saavedra, and S. Canuto, J. Mol. Struct.: THEOCHEM **466**, 69 (1999).
[50] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).
[51] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[52] *Handbook of Chemistry and Physics*, 73rd ed., edited by D. R. Lide (CRC Press, Boca Raton, FL, 1992).
[53] H. J. C. Berendsen, J. P. M. Postma, and W. F. van Gunsteren, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), p. 331.
[54] W. L. Jorgensen, J. M. Briggs, and M. L. Contreras, J. Phys. Chem. **94**, 1683 (1990).
[55] W. L. Jorgensen and N. A. McDonald, J. Mol. Struct.: THEOCHEM **424**, 145 (1998).
[56] S. Canuto and K. Coutinho, Adv. Quantum Chem. **28**, 90 (1997).
[57] K. Coutinho, M. J. de Oliveira, and S. Canuto, Int. J. Quantum Chem. **66**, 249 (1998).
[58] S. Canuto, K. Coutinho, and D. Trzesniak, Adv. Quantum Chem. (in press).
[59] K. Coutinho and S. Canuto, DICE (version 2.8): A general Monte Carlo program for liquid simulation, University of São Paulo, 2000.
[60] A. Rahman and F. H. Stilinger, J. Chem. Phys. **55**, 3336 (1971).
[61] F. H. Stilinger and A. Rahman, J. Chem. Phys. **60**, 1545 (1974).
[62] F. H. Stilinger, Adv. Chem. Phys. **31**, 1 (1975).
[63] M. Mezei and D. L. Beveridge, J. Chem. Phys. **74**, 622 (1981).
[64] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[65] C. Chatfield, *The Analysis of Time Series. An Introduction*, 3rd ed. (Chapman and Hall, New York, 1984).
[66] R. Krätschmer, K. Binder, and D. Stauffer, J. Stat. Phys. **15**, 267 (1976).
[67] S. Tang and D. P. Landau, Phys. Rev. B **36**, 567 (1987).
[68] H. Müller-Krumbhaar and K. Binder, J. Stat. Phys. **8**, 1 (1973)
[69] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 98, Revision A.6, Gaussian, Inc., Pittsburgh, PA, 1998.
[70] K. K. Innes, I. G. Ross, and W. R. Moomaw, J. Mol. Spectrosc. **132**, 492 (1988).
[71] W. S. Benedict, N. Gailer, and E. K. Plyler, J. Chem. Phys. **24**, 1139 (1956).
[72] J. D. Cruzan, L. B. Braly, K. Liu, M. G. Brown, J. G. Loeser, and R. J. Saykally, Science **271**, 59 (1996).
[73] S. S. Xantheas, J. Chem. Phys. **100**, 7523 (1994).
[74] L. Rincón, R. Almeida, D. García-Aldea, and H. D. Riega, J. Chem. Phys. **114**, 5552 (2001).