

# Combined Monte Carlo and quantum mechanics study of the hydration of the guanine-cytosine base pair

Kaline Coutinho,<sup>1</sup> Valdemir Ludwig,<sup>2</sup> and Sylvio Canuto<sup>2,\*</sup>

<sup>1</sup>Universidade de Mogi das Cruzes, CP 411, 08701-970, Mogi das Cruzes, São Paulo, Brazil

<sup>2</sup>Instituto de Física, Universidade de São Paulo, CP 66318, 05315-970, São Paulo, Brazil

(Received 18 December 2003; published 1 June 2004)

We present a computer simulation study of the hydration of the guanine-cytosine (GC) hydrogen-bonded complex. Using first principles density-functional theory, with gradient-corrected exchange-correlation and Monte Carlo simulation, we include thermal contribution, structural effects, solvent polarization, and the water-water and water-GC hydrogen bond interaction to show that the GC interaction in an aqueous environment is weakened to about 70% of the value obtained for an isolated complex. We also analyze in detail the preferred hydration sites of the GC pair and show that on the average it makes around five hydrogen bonds with water.

DOI: 10.1103/PhysRevE.69.061902

PACS number(s): 87.14.Gg., 82.30.Rs, 31.70.Dk, 31.15.Ar

## I. INTRODUCTION

The study of hydrogen bonds has been a central topic in chemistry and biology [1] for many years and is now increasingly recognized to be very important also in physics [2–5]. Hydrogen bonds between biomolecules is perhaps the most significant process regulating functions in living systems. Most of our understanding of the electronic structure of hydrogen bonds, and the associated properties and spectra, have been derived from studies in the crystalline phase or isolated complexes. An important step now is the understanding of hydrogen bonds in an explicit aqueous environment. Whereas in a hydrogen bonded complex the structure is fixed by a minimum-energy condition, in the liquid there are several configurations governed by the temperature and the related statistical disorder. Instead of a single value the electronic properties are more properly described by a statistical distribution of possible values within a certain width [6]. Understanding the differences of the hydrogen bonds formed between two organic molecules in an aqueous environment is crucial for describing several solvation and biomolecular processes [7]. The triple bond in guanine-cytosine (GC) has a binding energy that is estimated [8,9] to be 21 kcal/mol, whereas the double bond in adenine-thymine (AT) is estimated to be around 12 kcal/mol. These values correspond to gas phase calculations or thermodynamic gas phase measurements [8]. Extrapolating hydrogen bond information from the gas phase to the biological aqueous environment can be risky. Theoretical estimates have considered minimum-energy structure [10,11] instead of the intrinsic statistical nature of the liquid [6]. Contrary to AT, the stronger hydrogen-bonded GC pair is expected to be stable in water [12–14]. Hence the GC base pair (Fig. 1) provides a unique DNA base-pair system for the study of hydration of a hydrogen bond. In this paper we present an explicit, and statistically converged, *ab initio* determination of the change in the structure and interaction energy of the hydrated GC

pair in the Watson-Crick conformation. This is fundamental to understanding and rationalizing hydrogen bonds in the natural biological solvent. An additional point of considerable interest is the identification of the preferred hydration sites. It has been discussed that these hydration sites mark the position where protein residues hydrogen bond to DNA [15,16]. Therefore a complete statistical analysis of the hydrogen bonds formed between the GC base pair and water is presented. Although continuum models can be used to study solvent effects in hydrogen bonds [7], its use is not recommended here since one of the major concerns is the determination of the preferred hydration sites of the GC pair, which requires explicit consideration of the solvent water molecules. In this work the structure of the liquid is obtained from Monte Carlo (MC) simulation.

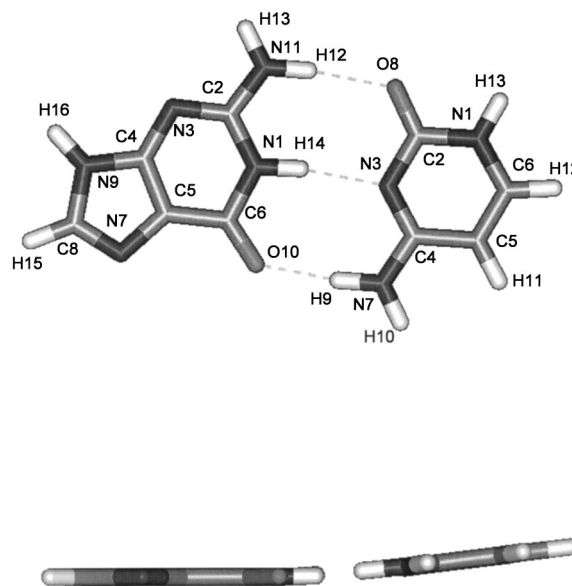


FIG. 1. (Color online) Guanine-cytosine base pair and definition of the indices. The bottom shows the lateral view with the interplanar angle of 7°.

\*Electronic address: canuto@if.usp.br

## II. THEORETICAL METHODOLOGY

The major difficulty in studying liquid systems is the realistic representation of the configuration space occupied by the molecules of the liquid. We use here a sequential procedure where MC simulations are made to generate the structure of the liquid and subsequently quantum mechanical (QM) calculations are performed on sampled structures. The sampling is carefully considered, aiming at obtaining statistically converged results with a small number of QM calculations.

To understand the effect of the aqueous environment in the triple hydrogen-bonded GC base pair, we calculated structural properties, interaction energy, and dipole moment of the base pair in three situations: (i) in the minimum-energy geometry of the isolated system (complex), (ii) in the configurations obtained in a Monte Carlo (MC) simulation of the isolated hydrogen-bonded GC system at  $T=298$  K (gas), and (iii) in the configurations obtained in a MC simulation of one GC pair embedded in as much as 1000 water molecules, also at  $T=298$  K (aqueous). The first, case (i), corresponds to the isolated optimized complex and can be compared to most previous calculations. The second, case (ii), includes only the relative thermal motion between the two base pairs corresponding to the base pairs in a gas, and the results can be compared with those obtained from gas phase measurements. Finally, the third, case (iii), is the one of most interest here, where the GC pair is completely surrounded by bulk water.

The first (gas) simulation used the Metropolis sampling in the canonical ( $NVT$ ) ensemble [17] and the second (aqueous) simulation used the isothermal-isobaric ( $NPT$ ) ensemble [17] at the pressure of 1 atm. In the gas simulation, standard procedures are used for the Metropolis sampling in the canonical ( $NVT$ ) ensemble using a cubic box with 35 Å of length. Each molecule was kept in its rigid geometry obtained in the complex optimization (see Table I), but left free to translate and rotate using the all-atom optimized parameters for the liquid simulation (OPLS-AA) potential [18] developed by Jorgensen and co-authors (shown in Table I). The simulation was started with the following complex configuration:  $2 \times 10^3$  MC steps were performed in the thermalization stage and  $4 \times 10^4$  MC steps in the equilibrium stage. In the aqueous simulation, standard procedures are also used for the Metropolis sampling in the isothermal-isobaric ( $NPT$ ) ensemble with pressure of 1 atm. The simulations started with a Hartree-Fock complex optimized configuration that was used only for the thermalization. After a thermalization stage of  $25 \times 10^6$  MC steps, the simulation consisted of  $15 \times 10^7$  MC steps for the production in equilibrium. A new MC step was obtained randomly, selecting a molecule and trying to translate it in all the Cartesian directions ( $\delta x$ ,  $\delta y$ , and  $\delta z$ , chosen randomly between  $-\delta r_{max}$  and  $+\delta r_{max}$ ) and also rotating it around a randomly chosen axis. As usual [17], the maximum allowed displacement  $\delta r_{max}$  was adjusted to give an acceptance rate around 50%, but the maximum rotation angle was fixed to  $\delta\theta = \pm 15^\circ$ . Each molecule ( $G$ ,  $C$ , and all water) was free to translate and rotate using the same all-atom OPLS potential developed by Jorgensen and co-authors [18] for the guanine and cytosine and the simple-

point charge (SPC) potential for water [19]. The intermolecular interactions are described by the standard Lennard-Jones-plus-Coulomb potential (see Table I):

$$U_{ab} = \sum_i^a \sum_j^b 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j e^2}{r_{ij}}. \quad (1)$$

In the calculation of the pair-wise energy, each molecule interacts with all other molecules within the cutoff radius that is half the size of the length of the cubic box. For larger separations we used the long-range correction of the potential energy [17]. For the Lennard-Jones potential, the long-range correction is evaluated using the pair radial distribution function, and for the Coulomb potential it is evaluated using the reaction field method with dipole interactions (dielectric constant of water=78.5).

An important point now is the adequate sampling of configurations for the subsequent quantum-mechanics calculations. It is well known that successive configurations generated in the MC simulations are statistically highly correlated, and will not give important additional information. Therefore we first calculate the interval of statistical correlation using the autocorrelation function of the energy [20]. The sampling of configurations from the MC simulations was made using the calculated interval of statistical correlation. We only select configurations that give relevant statistical information [20,21]. Here we used the calculated autocorrelation function [20] of the energy to extract 75 configurations, with less than 15% of statistical correlation. These configurations will be submitted to the QM calculations. As we shall see, this is enough to ensure statistical convergence.

The QM calculations were carried out with first-principle density-functional theory in the hybrid B3LYP gradient-corrected exchange correlation approximation [22] using localized basis functions in the B3LYP/6-31+ $G(d)$  method using the GAUSSIAN/98 program [23]. This theoretical model has been successfully used in previous calculations of base pairs [24,25]. All interaction (binding) energies are obtained using a counterpoise correction [26] to a basis set superposition error (BSSE), i.e.,

$$\Delta E_{A-B}^{CP} = E_{A-B} - E_{A(B)} - E_{B(A)}, \quad (2)$$

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 to the monomer  $B$ . This corresponds to a full counterpoise correction. All Monte Carlo simulations and classical-quantum interfaces were made with the DICE program [27].

## III. RESULTS AND DISCUSSIONS

Table II shows the results for the hydrogen bond distances of GC. These results were used only to start the MC simulations. As noted before [24,25], quantum chemical calculations of the minimum-energy structure give hydrogen bonds for the GC pair that is too short. The statistical average result obtained from the MC simulation gives an improved picture

TABLE I. Geometries and potential parameters of the Monte Carlo simulation. Cartesian coordinates in Å,  $q$  in elementary charge units,  $\epsilon$  in kcal/mol, and  $\sigma$  in Å. See text.

| Atoms    | Geometry          |         |         | Potential parameters |            |          |
|----------|-------------------|---------|---------|----------------------|------------|----------|
|          | $x$               | $y$     | $z$     | $q$                  | $\epsilon$ | $\sigma$ |
| Cytosine | HF/6-31G( $d,p$ ) |         |         | OPLS-AA              |            |          |
| N1       | 4.3007            | 0.9630  | 0.0872  | -0.560               | 0.170      | 3.250    |
| C2       | 2.9186            | 1.0444  | -0.0068 | 0.550                | 0.105      | 3.750    |
| N3       | 2.2292            | -0.1126 | -0.0596 | -0.540               | 0.170      | 3.250    |
| C4       | 2.8450            | -1.2731 | -0.0157 | 0.460                | 0.080      | 3.500    |
| C5       | 4.2845            | -1.3740 | 0.0842  | -0.060               | 0.080      | 3.500    |
| C6       | 4.9567            | -0.2197 | 0.1325  | 0.100                | 0.080      | 3.500    |
| N7       | 2.1121            | -2.3730 | -0.0661 | -0.790               | 0.170      | 3.250    |
| O8       | 2.4068            | 2.1405  | -0.0350 | -0.480               | 0.210      | 2.960    |
| H9       | 1.1092            | -2.2977 | -0.1118 | 0.385                | 0.000      | 0.000    |
| H10      | 2.5358            | -3.2682 | -0.0231 | 0.355                | 0.000      | 0.000    |
| H11      | 4.7820            | -2.3224 | 0.1199  | 0.100                | 0.050      | 2.500    |
| H12      | 6.0266            | -0.1697 | 0.2088  | 0.100                | 0.050      | 2.500    |
| H13      | 4.7859            | 1.8301  | 0.1256  | 0.380                | 0.000      | 0.000    |
| Guanine  | HF/6-31G( $d,p$ ) |         |         | OPLS-AA              |            |          |
| N1       | -0.7726           | 0.3302  | -0.0977 | -0.560               | 0.170      | 3.250    |
| C2       | -1.3836           | 1.5457  | -0.0498 | 0.460                | 0.080      | 3.500    |
| N3       | -2.6755           | 1.7009  | 0.0200  | -0.510               | 0.170      | 3.250    |
| C4       | -3.3364           | 0.5293  | 0.0389  | 0.340                | 0.080      | 3.500    |
| C5       | -2.8317           | -0.7460 | -0.0064 | 0.120                | 0.080      | 3.500    |
| C6       | -1.4165           | -0.9071 | -0.0807 | 0.520                | 0.105      | 3.750    |
| N7       | -3.8511           | -1.6766 | 0.0386  | -0.490               | 0.170      | 3.250    |
| C8       | -4.9195           | -0.9820 | 0.0107  | 0.200                | 0.080      | 3.500    |
| N9       | -4.6793           | 0.3741  | 0.1118  | -0.500               | 0.170      | 3.250    |
| O10      | -0.7666           | -1.9264 | -0.1254 | -0.510               | 0.210      | 2.960    |
| N11      | -0.5828           | 2.6155  | -0.0855 | -0.800               | 0.170      | 3.250    |
| H12      | 0.4143            | 2.5377  | -0.0775 | 0.400                | 0.000      | 0.000    |
| H13      | -1.0157           | 3.5041  | -0.0086 | 0.400                | 0.000      | 0.000    |
| H14      | 0.2332            | 0.2825  | -0.1310 | 0.380                | 0.000      | 0.000    |
| H15      | 0.4143            | -1.3725 | 0.1585  | 0.200                | 0.050      | 2.500    |
| H16      | -1.0157           | 1.1080  | 0.1592  | 0.350                | 0.000      | 0.000    |
| Water    | SPC               |         |         | SPC                  |            |          |
| O        | 0.0000            | 0.0000  | 0.0000  | -0.820               | 0.155      | 3.165    |
| H        | 0.5774            | -0.8165 | 0.0000  | 0.410                | 0.000      | 0.00     |
| H        | 0.5774            | 0.8165  | 0.0000  | 0.410                | 0.000      | 0.000    |

slightly increasing these distances. An important change compared to the gas phase results is an increase in the interplanar angle from the minimum-energy value of  $7^\circ$  to a larger average value of  $20 \pm 10^\circ$ . As a consequence of the spread, the interaction energy decreases. Table III shows that whereas the interaction energy in the minimum-energy configuration is calculated as 25.05 kcal/mol, in the gas, thermally averaged, this decreases to  $22.81 \pm 0.25$  kcal/mol. The experimental enthalpy [8] of 21 kcal/mol is in good agree-

ment with our calculated value. In the aqueous environment the hydrogen bond distances suffer some increase. The increase is small for the central N1-N3 distance (0.05 Å), but more pronounced and equivalent for the lateral. N11-O8(0.13 Å) and O10-N7(0.15 Å) distances. This is not a simple stretch of the bonds but an out-of-the-plane lateral swing of the relative positions of the molecules. The GC hydrogen bond interaction in water is found to be stable, i.e., no disruption and stacking have been obtained here. But the

TABLE II. Average distances (in Å) between GC obtained for the complex, the gas phase, and in an aqueous solution. The standard deviations are also shown. Atom labels are in Fig. 1.

| GC       | Complex <sup>a</sup> | Gas                 | Aqueous Solution    | Expt. <sup>b</sup> |
|----------|----------------------|---------------------|---------------------|--------------------|
| N11···O8 | 2.89±0.18            | 2.91±0.16           | 3.04±0.23           | 2.86               |
| N1···N3  | 2.91±0.16            | 2.91±0.09           | 2.96±0.13           | 2.95               |
| O10···N7 | 2.78±0.15            | 2.84±0.14           | 2.99±0.22           | 2.91               |
| $\theta$ | 7°                   | 20° (or -20°) ± 10° | 52° (or -52°) ± 20° |                    |

<sup>a</sup>Average over six *ab initio* geometry optimizations using different theoretical models.

<sup>b</sup>X-ray crystallographic measurements by Rosenberg *et al.* [28] on sodium guanylyl-3',5'-cytidine nonahydrate containing the Watson-Crick-type dimer (GpC)<sub>2</sub>. See also the theoretical result of Guerra *et al.* [24].

interplanar angle increases considerably to an average value of  $52 \pm 20^\circ$  with an increase in the width also. We also observed a few structures where the planes are perpendicular and also a limited number where they are reversed, showing only two hydrogen bonds H14-O8 and H12-N3, corresponding to the reverse Watson-Crick structure [1]. Figure 2 shows the superposition of the structures (only GC is shown to simplify the picture) obtained and the swing and increase in the interplanar angle. This is the configuration space occupied by cytosine as seen by the guanine molecule. As it can be noted there is a large increase in configuration space when the GC pair is in water.

Because the hydrogen bond acceptors of both guanine and cytosine are accessible to water, there is a great interest in establishing whether there are intruder water molecules that bind to either cytosine or guanine. In fact, it has been proposed that protein atoms involved in binding to DNA occupy positions normally occupied by water molecules in unbound DNA [15,16]. Thus the localization of the preferred hydration sites is of great general interest. Schneider and Berman [29] considered this for the crystalline case. The position of hydration sites were obtained from pseudoelectron densities and give an indication of the highest probability to find water molecules that are hydrogen bonded to the GC pair in the crystalline phase. To obtain this result for the liquid case we extracted 3000 uncorrelated configurations of the MC simulation of GC in water. Using a geometric and energetic criterion [6,20], and the radial distribution functions, we have determined the average number of hydrogen bonds in every atomic site of the GC pair in liquid water. The results are shown in Table IV. The oxygen atom of the guanine molecule is the one with more hydrogen bonds, with an average of 1.5. In fact this is not much different for the oxygen atom

of cytosine, where we find an average of 1.2 hydrogen bonded water molecules. On the other hand, the N3 atom of cytosine is involved in the GC bond but still makes an average of 0.3 hydrogen bonds with water. These bridging water molecules that attach in the hydrogen bonded atoms of the GC pair influence the base pair binding and were also seen in a recent theoretical study [30]. In total, there are an average number of 4.7 hydrogen bonds made between water and the GC pair. This is in agreement with the results of Schneider and Bermann [29], that suggested a number between 3 and 5. In particular, the preferred hydration sites of guanine (N7, O10, and N3) and cytosine (O8, N7) are in agreement with the average crystallographic and optimized structure [11]. Figure 3 shows the complete statistical distribution of values. As can be seen, 52% of the configurations have four or five hydrogen bonds between GC and water. As an illustration, Fig. 4 shows one of these configurations with five water molecules that are hydrogen-bonded to the GC base pair. The distribution of Fig. 3 also shows that there are a (small) number of configurations that can make eight or even nine hydrogen bonds. This indicates very clearly that one single configuration cannot represent the structure of the hydrogen bond hydration shell in the liquid. The recent interest in the influence of this hydration shell led to studies of geometry-optimized complexes of GC (and similar systems) and a few water molecules [11,31,32]. These studies required eight water molecules in a complex-optimized structure to obtain a sizable decrease in the interaction energy and increase in the interplanar angle [31]. In the liquid, configurations with eight water molecules in the hydration shell are statistically negligible, representing only 1.5% of the total configurations (Fig. 3). This result is indicative of the great importance of properly considering the statistical distribution of hydrogen bonded water molecules.

TABLE III. Interaction energies (in kcal/mol) and dipole moment (in Debye) for the GC pair. All results obtained with B3LYP/6-31+G(d). The standard deviations are also shown.

| Method               | Pristine <sup>b</sup> |      | Complex | Gas        | Aqueous solution |            |
|----------------------|-----------------------|------|---------|------------|------------------|------------|
|                      | G                     | C    |         |            | GC               | GC         |
| Binding <sup>a</sup> |                       |      | 25.05   | 22.81±0.25 | 17.17±0.55       | 17.32±0.65 |
| Dipole               | 7.17                  | 7.04 | 6.01    | 6.42±0.11  | 7.18±0.17        | 7.95±0.20  |

<sup>a</sup>The experimental [8] gas phase value is 21.0 kcal/mol.

<sup>b</sup>Includes counterpoise correction.

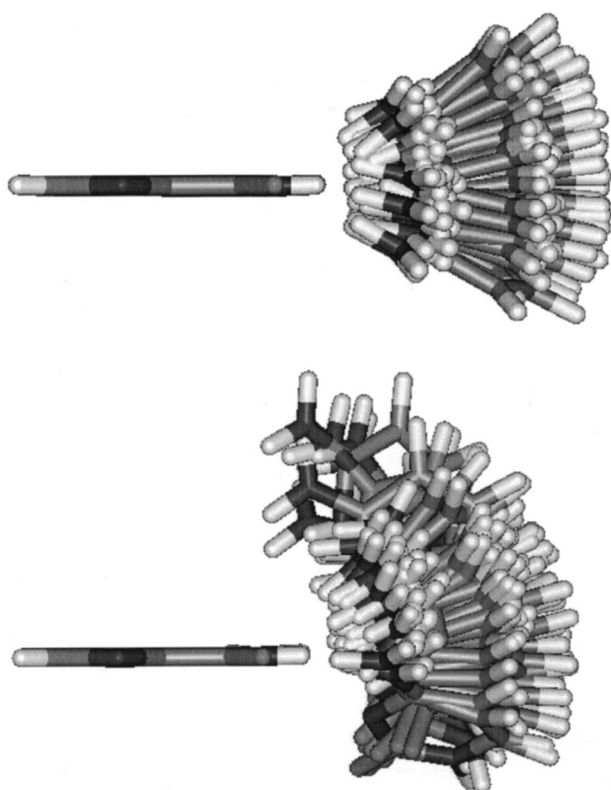


FIG. 2. (Color online) Superposition of 75 configurations of the GC pair obtained from the Monte Carlo simulation in gas phase (top) and in aqueous solution (bottom).

Now we consider the interaction energy of the GC base pair in the aqueous environment. The statistically relevant configurations obtained from the MC simulation are extracted for the B3LYP/6-31+ $G(d)$  calculations of the energy of the pair and the monomers [Eq. (2)] to obtain the average value. Table III shows that the GC interaction energy is now calculated as  $17.17 \pm 0.55$  kcal/mol. This decrease has most of its origin in the statistical distribution of possible configurations and also in the network of hydrogen bonds that influences the interaction of the GC pair. The influence of the hydrogen bonded water molecules in the binding energy of the GC pair has been studied recently using the positions of the water molecules in local energy minima [30–32]. This present contribution emphasizes instead the importance of considering the statistical nature of the configurations of hydrogen bonded water molecules. The considerations of the outer water molecules, beyond the hydrogen bond shell, are also important in order to understand the influence of the network of the hydrogen-bonded water molecules in the GC interaction (see Fig. 5). Whereas in the isolated case they are free to assume a minimum-energy configuration in the aque-

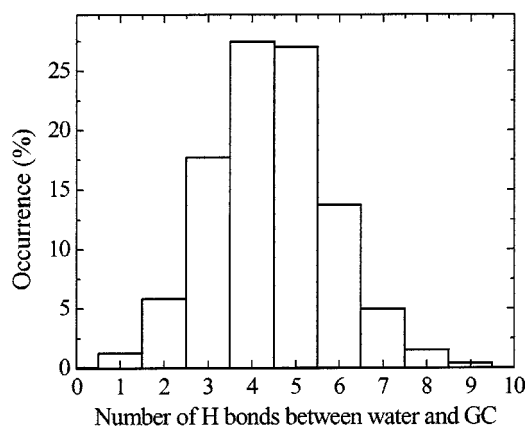


FIG. 3. Histogram of the number of hydrogen bonds formed between the GC pair and the solvent water molecules obtained from the Monte Carlo simulation.

ous environment, aside from the thermal contribution, the water molecules stick to the GC pair leading to both a large spread in the possible configurations and also holding the direct interaction between the pair.

An important point now is the consideration of the statistical convergence of the interaction energy. Figure 6 shows that the average binding energies are indeed statistically converged. In the thermal gas phase there is a faster convergence with a smaller statistical error because of the small width of calculated values. We obtain converged results already with  $\sim 30$  QM calculations of the binding energy. This is because only statistically relevant configurations are used. This is one important advantage of performing first the classical simulation. All statistically relevant information is available before running into the costly QM calculations.

Before coming to the final estimate of the interaction energy in solution we should include now the electrostatic field of the solvent water molecules. The water molecules are represented only by their simple-point charges (SPC) explicitly included in the QM calculations. Therefore, the QM calculations are performed now for the GC pairs and all (345) SPC water molecules within the radius of  $13.3 \text{ \AA}$  from the center of mass of the GC pair. This corresponds to the GC pair and all water molecules up to the third solvation shell in the center-of-mass radial distribution function. The results are shown in the last column of Table III (GC+345  $\text{H}_2\text{O}_{\text{SPC}}$ ) and converge to the interaction energy of  $17.32 \pm 0.65$  kcal/mol, corresponding to a slight increase of  $0.15$  kcal/mol for the interaction in the aqueous environment. This difference is smaller than the standard deviation of  $0.65$  kcal/mol, showing that it is not statistically significant and that the electrostatic field of the solvent water is essentially negligible. This is our best estimate of the binding energy, corresponding to

TABLE IV. Statistics of the hydrogen bonds between GC and proton-donor water (labels in Fig. 1).

|               | Guanine |     |     |     | Cytosine |     |     | Total |
|---------------|---------|-----|-----|-----|----------|-----|-----|-------|
| Atom label    | N3      | N7  | O10 | N11 | N3       | N7  | O8  |       |
| Average of HB | 0.7     | 0.8 | 1.5 | 0.1 | 0.3      | 0.1 | 1.2 | 4.7   |

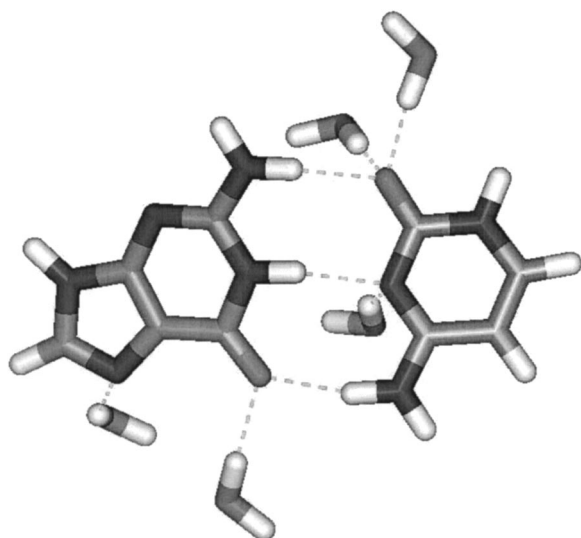


FIG. 4. (Color online) Illustration showing one typical configuration of the Monte Carlo simulation, with five water molecules hydrogen-bonded to the GC pair.

approximately 70% of the energy in the minimum-energy structure. The electrostatic interaction with the water molecules adds little to the binding energy, provided the calculations are performed on the GC pair using the configurations of the aqueous simulation.

In addition, we analyzed the electrostatic contribution to the polarization of the GC pair. Table III shows that the dipole moment of the base pair increases from 6.01 D to  $7.95 \pm 0.20$  D in the aqueous medium, with approximately 10% being due to the electrostatic polarization coming from the aqueous environment. Figure 7 shows the calculated histogram and the converged best theoretical value of  $7.95 \pm 0.20$  D.

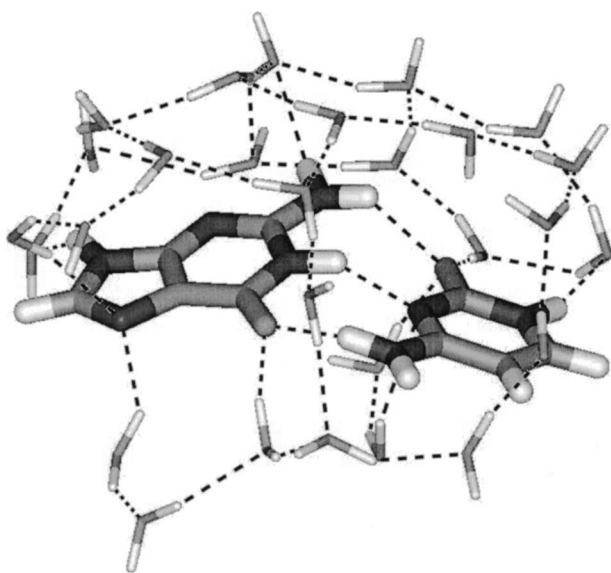


FIG. 5. (Color online) Illustration of the network of nearest hydrogen-bonded water molecules around the GC pair, obtained from the Monte Carlo simulations.

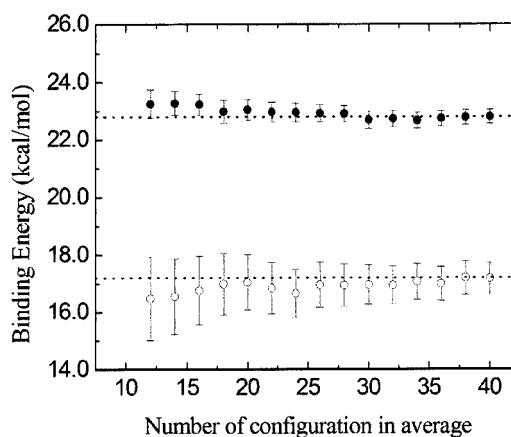


FIG. 6. Convergence of the calculated average binding energy with the number of Monte Carlo configurations. The error bar is the statistical error. The top is obtained using the simulation for the isolated GC pair (gas) and the bottom is for the GC pair in the aqueous environment. See the text and Table III.

#### IV. SUMMARY AND CONCLUSION

The interaction between the base pairs of the DNA has been subjected to numerous theoretical calculations. The next step is the understanding of the role of the aqueous environment in the base pair interaction and conformation. The proper treatment of the liquid requires statistical mechanics. This research is a step in this direction. As the preferred hydration sites are associated with possible marks for the binding of protein to DNA, there has been considerable interest in the precise determination of the hydrogen bonded water. Most of the previous studies have considered minimum-energy configurations, thereby missing the important statistical aspect. In this study we have made a detailed analysis of the hydrogen bonds between proton-donor water and the guanine-cytosine base pair. We identify the preferred sites and find that on the average there are 4.7 water molecules that are hydrogen-bonded to the base pair. This average and the most statistically relevant sites are in close agreement with crystallographic and previous theoretical data [11,29]. Considering the guanine-cytosine pair in the aque-

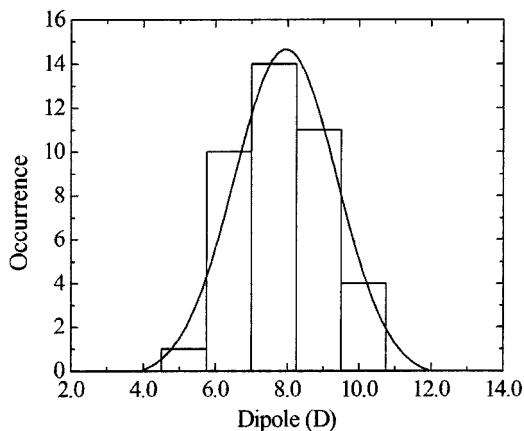


FIG. 7. Histogram of the calculated dipole moment for the Monte Carlo configurations of the GC pair in water.

ous environment at room temperature, we obtain from first-principle calculations and statistically converged results that the interaction energy is reduced to  $\sim 2/3$  of the value in the isolated complex. The interaction energy obtained here includes the thermodynamic disorder and the influence of the configuration space that is occupied by the base pair in water. This is, of course, not the same as the free energy of formation in solution that is estimated to be a much lower value, around 6 kcal/mol, in the case of this base pair [33]. The strength of an enzymatic inhibitor is the starting point for modeling and rationalizing its action in the active site. As most reversible inhibitors are in fact hydrogen bonded to the

aminoacid of the active site, this example emphasizes the importance of obtaining reliable values for the interaction energy of hydrated hydrogen bonds. Crystalline or gas phase results cannot be extrapolated to the aqueous environment without great risk and implications in many biomolecular processes.

#### ACKNOWLEDGMENTS

This work is partially supported by CNPq and FAPESP. We thank CENAPAD-SP and LCCA for computational support.

- 
- [1] G. A. Jeffrey, *An Introduction to Hydrogen Bond* (Oxford University, London, 1997).
- [2] K. R. Wilson, J. G. Tobin, A. L. Ankudinov, J. J. Rehr, and R. J. Saykally, *Phys. Rev. Lett.* **85**, 4289 (2000).
- [3] M. Boero, K. Terakura, T. Ikeshoji, C. C. Liew, and M. Parrinello, *Phys. Rev. Lett.* **85**, 3245 (2000).
- [4] K. Modig, B. Pfrommer, and B. Halle, *Phys. Rev. Lett.* **90**, 075502 (2003).
- [5] J.-H. Guo, Y. Luo, A. Augustsson, S. Kashtanov, J.-E. Rubensson, D. K. Shuh, H. Ågren, and J. Nordgren, *Phys. Rev. Lett.* **91**, 157401 (2003).
- [6] E. E. Fileti, K. Coutinho, T. Malaspina, and S. Canuto, *Phys. Rev. E* **67**, 061504 (2003).
- [7] A. J. A. Aquino, D. Tunega, G. Haberhauer, M. H. Gerzabek, and H. Lischka, *J. Phys. Chem. A* **106**, 1862 (2002).
- [8] I. K. Yanson, A. B. Teplitsky, and L. F. Sukhodub, *Biopolymers* **18**, 1149 (1979).
- [9] J. Sponer, J. Leszczynski, and P. Hobza, *J. Phys. Chem.* **100**, 1965 (1996).
- [10] M. Kabelác and P. Hobza, *J. Phys. Chem. B* **105**, 5804 (2001).
- [11] F. Moroni, A. Famulari, and M. Raimondi, *J. Phys. Chem. A* **105**, 1169 (2001).
- [12] M. Kabelác, F. Ryjáček, and P. Hobza, *Phys. Chem. Chem. Phys.* **2**, 4906 (2000).
- [13] J. Florián, J. Sponer, and A. Warshel, *J. Phys. Chem. B* **103**, 884 (1999).
- [14] V. I. Danilov, N. V. Zheltovsky, O. N. Slyusarchuk, V. I. Poltev, and J. L. Alderfer, *J. Biomol. Struct. Dyn.* **15**, 69 (1997).
- [15] N. C. Seeman, J. M. Rosenberg, and A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 804 (1976).
- [16] J. Woda, B. Schneider, K. Patel, K. Mistry, and H. M. Berman, *Biophys. J.* **75**, 2170 (1998).
- [17] M. P. Allen, and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, U.K., 1987).
- [18] J. Pranata, S. G. Wierschke, and W. L. Jorgensen, *J. Am. Chem. Soc.* **113**, 2810 (1991).
- [19] H. J. C. Berendsen, J. P. M. Postma, and W. F. van Gunsteren, *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), p. 331.
- [20] K. Coutinho, S. Canuto, and M. C. Zerner, *J. Chem. Phys.* **112**, 9874 (2000).
- [21] T. Malaspina, K. Coutinho, and S. Canuto, *J. Chem. Phys.* **117**, 1692 (2002).
- [22] A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993); C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1998).
- [23] M. J. Frisch *et al.*, GAUSSIAN-98, Gaussian Inc., Pittsburgh, PA, 1998.
- [24] C. F. Guerra, F. M. Bickelhaupt, J. G. Snijders, and E. J. Baerends, *J. Am. Chem. Soc.* **122**, 4117 (2000).
- [25] R. Rivelino, V. Ludwig, E. Rissi, and S. Canuto, *J. Mol. Struct.* **615**, 257 (2002).
- [26] S. F. Boys and F. Bernardi, *Mol. Phys.* **19**, 553 (1970).
- [27] K. Coutinho, and S. Canuto, DICE (version 2.8): *A General Monte Carlo Program for Liquid Simulation* (University of São Paulo, São Paulo, Brazil, 2002).
- [28] J. M. Rosenberg, N. C. Seeman, R. O. Day, and A. Rich, *J. Mol. Biol.* **104**, 145 (1976).
- [29] B. Schneider and H. M. Berman, *Biophys. J.* **69**, 2661 (1995).
- [30] E. González, A. Deriabina, A. Teplukhin, A. Hernández, and V. I. Poltev, *Theor. Chem. Acc.* **110**, 460 (2003).
- [31] N. U. Zhanpeisov and J. Leszczynski, *J. Mol. Struct.: THEOCHEM* **487**, 107 (1999).
- [32] N. U. Zhanpeisov, J. Sponer, and J. Leszczynski, *J. Phys. Chem. A* **102**, 10374 (1998).
- [33] E. Stofer, C. Chipot, and R. Lavery, *J. Am. Chem. Soc.* **121**, 9503 (1999).