Study of the optical and magnetic properties of pyrimidine in water combining PCM and QM/MM methodologies

Vinicius Manzoni, Marcelo L. Lyra, Rodrigo M. Gester, Kaline Coutinho and Sylvio Canuto

Received 4th April 2010, Accepted 17th August 2010
DOI: 10.1039/c0cp00122h

The solvent effects on the low-lying absorption spectrum and on the $^{15}$N chemical shielding of pyrimidine in water are calculated using the combined and sequential Monte Carlo simulation and quantum mechanical calculations. Special attention is devoted to the solute polarization. This is included by an iterative procedure previously developed where the solute is electrostatically equilibrated with the solvent. In addition, we verify the simple yet unexplored alternative of combining the polarizable continuum model (PCM) and the hybrid QM/MM method. We use PCM to obtain the average solute polarization and include this in the MM part of the sequential QM/MM methodology, PCM-MM/QM. These procedures are compared and further used in the discrete and the explicit solvent models. The use of the PCM polarization implemented in the MM part seems to generate a very good description of the average solute polarization leading to very good results for the n–π* excitation energy and the $^{15}$N nuclear chemical shield of pyrimidine in aqueous environment. The best results obtained here using the solute pyrimidine surrounded by 28 explicit water molecules embedded in the electrostatic field of the remaining 472 molecules give the statistically converged values for the low lying n–π* absorption transition in water of 36 900 ± 100 (PCM polarization) and 36 950 ± 100 cm$^{-1}$ (iterative polarization), in excellent agreement among one another and with the experimental value observed with a band maximum at 36 900 cm$^{-1}$. For the nuclear shielding $^{15}$N the corresponding gas–water chemical shift obtained using the solute pyrimidine surrounded by 9 explicit water molecules embedded in the electrostatic field of the remaining 491 molecules give the statistically converged values of 24.4 ± 0.8 and 28.5 ± 0.8 ppm, compared with the inferred experimental value of 19 ± 2 ppm. Considering the simplicity of the PCM over the iterative polarization this is an important aspect and the computational savings point to the possibility of dealing with larger solute molecules. This PCM-MM/QM approach reconciles the simplicity of the PCM model with the reliability of the combined QM/MM approaches.

1. Introduction

In the last decades quantum chemistry has proved to be a reliable and useful tool to describe the electronic properties of isolated small and medium-sized molecules. As most chemical reactions and spectroscopic measurements occur in solution a natural extension has been the development of methods for including the influence of the liquid environment on the molecular properties and we have witnessed an enormous progress in this direction. The study of how the solvent can influence molecular properties in general has been one of the central concerns in quantum chemistry and molecular physics. There are essentially three dominant methodologies that involve different degrees of sophistication: the continuum, the discrete approach and the explicit use of solvent molecules. The continuum method is based on the original ideas of Kirkwood and Onsager where the medium is represented by its macroscopic dielectric constant. Several developments were made based on this idea and these include the self-consistent reaction field, the polarizable continuum model (PCM), the conductor-like screening model and the discrete reaction field, to quote some of the most used. An alternative that is gaining wide applicability is to consider the solvent at a microscopic level of description. The discrete and explicit models are inter-related and the first case could be considered as a simple approximation to the second. Both in the discrete (solvent treated only as point charges) and in the explicit models the location of the solvent molecules is important. The most systematic way for positioning the solvent molecules is by using some sort of computer simulation of liquids, performed in the corresponding thermodynamic condition. An efficient procedure samples statistically uncorrelated liquid structures for the quantum mechanical calculations. This is normally termed as a sequential quantum mechanics molecular mechanics methodology. The generation of the liquid structure can then be made using molecular dynamics or Monte Carlo simulation. Based on computational aspects, normally related to the cost, the computer simulation of liquids has been limited.

$^{a}$Instituto de Física, Universidade Federal de Alagoas, 57072-970, Maceió, AL, Brazil
$^{b}$Instituto de Física, Universidade de São Paulo, CP 66318, 05314-970, São Paulo, SP, Brazil. E-mail: canuto@if.usp.br
$^{c}$Departamento de Ciências Exatas e Naturais, Universidade Federal do Pará, 68505-080, Marabá, PA, Brazil
to classical Monte Carlo or Molecular Dynamics simulations. In specific cases, however, quantum mechanical simulations to generate the liquid structure are also possible and recently have been applied in the sequential QM/MM. For simple configuration averages we have favored classical Monte Carlo (MC) simulation. In this case quantum-mechanical calculations are then performed on solute–solvent configurations generated by MC simulation to obtain the statistically converged values of the molecular property of interest. The sequential (thus separated) use of Monte Carlo (MC) and Quantum Mechanics (QM) approaches is a flexible procedure and has the advantage that it can ensure systematic statistical convergence of the average results. But it has the disadvantage that the classical simulation is uncoupled from the (subsequent) QM calculations. This imposes the necessity to carefully analyze the mutual polarization between the solute and the solvent, known to be important. In most solvation studies one reference molecule is immersed in a solvent and for this extreme dilution the solvent polarization by the solute is relatively important. But much less compared to the solute polarization by the solvent. In these cases an alternative has been developed where the solute determination is used using an iterative procedure that aims at bringing the solute to an electrostatic equilibrium with the solvent. This requires a series of simulations followed by QM calculations. A simple but very useful polarization scheme is given by the PCM model, normally treated as intrinsically different approaches—PCM and QM/MM methods—we consider here a possible reconciliation. They can complement one another. We verify the simple yet unexplored alternative of combining the PCM and QM/MM methods. We thus use PCM to obtain the solute polarization and include this in the MM part of the sequential QM/MM methodology. We shall term this as PCM-MM/QM. Given its characteristics it is conceivable that the PCM method gives a good solute polarization and we explore this possibility.

We focus on the analysis of the low-lying n–π* absorption bands and the 15N nuclear magnetic shielding of pyrimidine in water. Understanding the effects of the aqueous environment on the spectroscopic properties of pyrimidine is important for several reasons. These include its role as a model for bio-molecular systems and because of the important contribution of the solute–solvent hydrogen bonds.

We analyze the absorption spectrum and the nuclear magnetic resonance (NMR) shielding of pyrimidine in water using different degrees of solvation and different solute polarizations. Thus we first use the continuum model per se to obtain the pyrimidine properties. Second, we use the solute polarization obtained by the PCM calculation to generate the atomic charges for the MC simulation. Third, we use the iterative polarization to similarly obtain the atomic charges for the MC simulation. In these two latter cases we compare the in-water dipole moment of pyrimidine and the subsequent calculated values of the pyrimidine spectroscopic properties in water. In addition, in the two different polarizations (PCM and iterative procedures) we analyze the role of using only point charges to represent the solvent (discrete model) and the inclusion of some explicit solvent molecules. As stated before we analyze the absorption spectrum and the 15N nuclear magnetic shielding parameters. It is expected that NMR parameters are more affected by local solvent effects whereas the excitation energies may also be influenced by long-range effects. Thus we examine two complementary properties. As pyrimidine is an acceptor of hydrogen bonds and because water is a very protic solvent, when examining the NMR parameters, we pay special attention to the role of the solute–solvent hydrogen bonds formed in both nitrogen sites.

2. Calculations details

We use the sequential Monte Carlo/quantum mechanics methodology to calculate the optical and magnetic properties of pyrimidine in water. Similar to previous studies we first generate the liquid configurations using MC simulation. Next, we perform QM calculations on these configurations, treated as supermolecules, composed of one pyrimidine surrounded by several water molecules.

The simulations were carried out with the program DICE in the isothermal–isobaric, NPT, ensemble in the Metropolis sampling technique using the image method and periodic boundary conditions. We have used the temperature of 298 K and pressure of 1 atm. The system is composed of one pyrimidine molecule surrounded by 1000 water molecules. The geometry of pyrimidine was obtained from a MP2/6-31+G(d,p) optimization. Intermolecular interactions were modeled by the standard Lennard-Jones (LJ) and Coulomb potential with three parameters for each interacting site. For water we have used the SPC potential. For pyrimidine we have used the LJ parameters of the OPLS force field and the charges were obtained using different procedures corresponding to the solute polarization model adopted. This will be discussed separately in Section 3. The simulation consists of two parts. In the thermalization, a total of \(2.1 \times 10^5\) steps were used and then followed by \(5 \times 10^5\) steps in the production stage. As it will be seen several MC simulations are made but in every case the configurations sampled for the subsequent QM calculations have less than 12% of statistical correlation, and in every case a total of 100 configurations are sampled for the QM calculations as done before.
The excitation energies are obtained using the time dependent density functional theory (TDDFT)\textsuperscript{51} with the three-parameter Becke exchange functional\textsuperscript{52} and the Lee–Yang–Parr correlation\textsuperscript{53} with a basis set that includes both diffuse and polarization functions, B3LYP/6-311 +G(d). Referring to the $^{15}$N magnetic shielding we treat the two nitrogen atoms as similar. For the calculations we use the same DFT functional but a basis set that is specially designed for obtaining nuclear magnetic shielding.\textsuperscript{54} For the target N atom we use a larger basis set, the aug-ccS2, and for the remaining atoms (C, O and H) we use pcS2. Thus the NMR parameter is calculated using B3LYP/aug-ccS2(N)/pcS2(C,O,H).

In the discrete model, the solvent is represented by only the electrostatic embedding (solvent represented by point charges only), thus instead of performing the QM calculations for each of the 100 configurations we have used the average solvent electrostatic configuration (ASEC)\textsuperscript{55} and have performed only one QM calculation. This ASEC procedure generates an average configuration following very closely the idea of the average configuration following very closely the idea of the average solvent electrostatic potential (ASEP) developed by Sánchez et al.\textsuperscript{56–58}

All PCM calculations used the integral equation formalism (IEF),\textsuperscript{43,44} all calculations of the NMR shielding constants have been made using the gauge-including atomic orbitals (GIAO)\textsuperscript{59,60} and all QM calculations used the Gaussian 03 package.\textsuperscript{61}

3. Results and discussion

3.1 Electronic polarization of pyrimidine in water

We now discuss the inclusion of the solute polarization effects. The original OPLS force field for pyrimidine\textsuperscript{50} has been parametrized to reproduce the thermodynamic properties of the homogeneous liquid. This includes atomic charges obtained\textsuperscript{50v} at the HF/6-31G(d) level using a CHELPG electrostatic mapping.\textsuperscript{62} Although the LJ parameters have been considered to be transferable the atomic charges have to be adapted to the pyrimidine–water interaction. The solute polarization adopted here thus adjusts the charges to the specific solvent used. Two procedures are used and the corresponding solute dipole moments will be compared. The PCM (PCM-MM/QM) and the iterative procedures (I-QM/MM) are then used to calculate the atomic charges of the solute molecule in the presence of the solvent. In the I-QM/MM procedure we first perform a MP2/aug-cc-pVQZ calculation for the isolated solute molecule and obtain the atomic charges using the CHELPG mapping.\textsuperscript{62} These are the charges initially used in the classical potential of the MC simulation. After this first MC simulation statistically uncorrelated configurations\textsuperscript{33,34} of the liquid are selected to calculate the average atomic charges and the dipole moment of the solute embedded in the electrostatic field of the solvent environment using MP2/aug-cc-pVQZ. These average atomic charges of the solute are updated in the Coulomb part of the potential for another MC simulation and subsequently another average atomic charges and dipole moment of the embedded solute are calculated. This process is repeated until a convergence in the QM average dipole moment is obtained, when the solute is then in electrostatic equilibrium with the solvent. These results are summarized in Fig. 2 that shows that the converged in-water average dipole moment of pyrimidine is obtained in five to seven iterations. Excluding the first point (iteration 0) every entry in this figure corresponds to a statistically converged average value of the in-water dipole moment of pyrimidine. The ASEC was generated with 100 statistically uncorrelated MC configurations of the nearest 500 water molecules represented by point charges. This includes all water molecules within the center of mass distance of 15.3 Å from pyrimidine. The dipole moment of the original all-atom OPLS force field is 2.52 D,\textsuperscript{50a} an improvement with respect to the united-atom value\textsuperscript{56b} of 2.16 D. In the gas phase (iteration 0), the calculated dipole moment of pyrimidine is 2.42 D. This should be compared with the value of 2.33 D for the experimental gas phase result.\textsuperscript{63} In turn, the converged in-water average value obtained here is 4.08 D, corresponding to an increase of ca. 70%. These in-water values are larger than a simple previous theoretical prediction of 2.93 D.\textsuperscript{64} For comparison, the corresponding in-water dipole moment obtained using the PCM model is 3.50 D. This gives the other alternative considered here that could show considerable computational savings; namely, to use directly the atomic charges obtained from the PCM calculation in a single MC simulation.

In this PCM polarization the atomic charges are used in the MC simulation to generate the solute–solvent configurations. This is not electrostatically equilibrated and the CHELPG electrostatic fitting, using again the same MP2/aug-cc-pVQZ, gives an increased dipole moment value of 3.89 D. All these dipole moments are shown in Fig. 2. The optical and magnetic properties of pyrimidine will be obtained using these two polarization schemes. The corresponding atomic charges of the final MC simulation that will generate the configuration used for the QM calculations are also shown in Table 1.

In-solution dipole moments are not easily obtained from experiments and only indirect results can be obtained. One possibility is by using experimental infrared intensities.\textsuperscript{65} For pyrimidine in water the value has apparently not been estimated experimentally.

All results that will be discussed in the following sections have used the polarized in-water pyrimidine with the atomic charges presented in Table 1.
### Table 1

<table>
<thead>
<tr>
<th>Site</th>
<th>( \varphi_{\text{PCM}} )</th>
<th>( \varphi_{\text{iterative}} )</th>
<th>( \varepsilon_i ) (kcal mol(^{-1}))</th>
<th>( \sigma_i ) (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C(<em>{4})N(</em>{3})H(_{4})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N(1)</td>
<td>-0.868</td>
<td>-0.977</td>
<td>0.170</td>
<td>3.250</td>
</tr>
<tr>
<td>N(2)</td>
<td>-0.868</td>
<td>-0.985</td>
<td>0.170</td>
<td>3.250</td>
</tr>
<tr>
<td>C(1)</td>
<td>0.866</td>
<td>0.986</td>
<td>0.110</td>
<td>3.750</td>
</tr>
<tr>
<td>C(2)</td>
<td>0.640</td>
<td>0.712</td>
<td>0.110</td>
<td>3.750</td>
</tr>
<tr>
<td>C(3)</td>
<td>-0.599</td>
<td>-0.618</td>
<td>0.110</td>
<td>3.750</td>
</tr>
<tr>
<td>C(4)</td>
<td>0.640</td>
<td>0.720</td>
<td>0.110</td>
<td>3.750</td>
</tr>
<tr>
<td>H(1)</td>
<td>-0.039</td>
<td>-0.059</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>H(2)</td>
<td>0.007</td>
<td>-0.003</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>H(3)</td>
<td>0.214</td>
<td>0.220</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>H(4)</td>
<td>0.007</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Dipole moment

\( \mu \) 3.50 + 1 MC 3.89 4.08

### 3.2 Structural aspects and solute–solvent hydrogen bonds

An important point in the study of computer simulation of liquids is the characterization of the structure of the solvent around the solute molecule. This signifies the proper statistical identification of the solvation shells and the solute–solvent hydrogen bonds. We thus analyze the pairwise radial distribution function because it gives the coordination number and defines the solvation shells. \(^{48}\) Fig. 3 shows the distribution function between the center of mass of the solute pyrimidine and the solvent water \( G_{\text{cm-cm}}(r) \) using the two polarization schemes adopted here.

The results are similar and they both show pronounced peaks around 4.0 and 7.5 Å. In fact at this level no structural differences are discernible between the PCM and the iterative polarization procedures, except for the details of the first peak. As it can be seen, the first hydration shell starts at 2.95 Å and ends at 5.95 Å where a second hydration shell starts and then ends at 9.2 Å. Spherical integration of these peaks in the \( G_{\text{cm-cm}}(r) \) gives a coordination number of 28 and 108 for these two shells respectively. At the first maximum located close to 4 Å a total of 9 water molecules are found. In the QM calculations we will consider separately the influence of using the discrete or the explicit solvent model. In the discrete case, all water molecules within the limit of 15.3 Å (500 water molecules) are considered but treated only as a simple point charge. In the explicit case, we consider separately 9 water molecules and the full first hydration shell, composed of 28 solvent water molecules. For both cases 100 QM calculations will be performed on supermolecular structures composed of one pyrimidine and 9 water and also the 28 water molecules of the first hydration shell. Fig. 4 illustrates one of these structures composed of 28 explicit water molecules and the remaining 472 treated as an electrostatic embedding. This signifies a QM problem involving 322 electrons in a wave function that is anti-symmetric with respect to all electrons, irrespective of whether it is in the solute or in the solvent. To obtain the average value 100 QM calculations are performed.

Now we consider the hydrogen bonds formed between the solute and the solvent. This is better characterized by analyzing the radial distribution function between the nitrogen atoms of pyrimidine and the hydrogen atoms of water \( G_{N-H}(r) \). Because of the difficulties in observing hydrogen atoms in X-ray experiments it is also convenient to study in complement the radial distribution function between the nitrogen atoms of pyrimidine and the oxygen atom of water \( G_{N-O}(r) \). Fig. 5 shows both \( G_{N-O}(r) \) and \( G_{N-H}(r) \) for the two polarizations considered. The \( G_{N-H}(r) \) distribution functions are similar to those obtained by Kongsted and Mennucci\(^{66}\) and others\(^{64,67-69}\) and they all show the pronounced hydrogen bond peak centered around 2 Å. This peak maximum is located at 1.96 Å in ref. 66 and at 1.85 Å here, in both polarizations.

This is in very good agreement with the experimental result of 1.82 Å derived from NMR experiments.\(^{70}\) Some small differences can be seen comparing the two polarization methods, but they are mostly associated with the peak intensities. Looking at the \( G_{N-H}(r) \) distribution we note that the iterative polarization gives a higher peak suggesting a larger number of hydrogen bonds. Both procedures agree that the hydrogen bonds are found within \( R_{N-O} = 3.2 \) Å. This is still not enough to identify the hydrogen bonds because it cannot be assured that all nearest-neighbor structures involved within a distance N–O smaller than the first minimum of \( G_{N-O}(r) \) (3.2 Å in this case) are indeed associated with hydrogen bonds. Combining this with the N–H distribution gives better results but structures not associated to hydrogen bond may still persist. A more efficient and correct way to extract the hydrogen-bonded

---

**Fig. 3** Radial distribution functions between the center of mass of pyrimidine and water.

**Fig. 4** Illustration of one configuration showing the pyrimidine surrounded by 28 water molecules and embedded in the electrostatic field of the remaining solvent molecules.
structures was discussed previously by Stillinger and Rahman, Mezei and Beveridge, Rapaport and frequently used by us. We identify a solute–solvent hydrogen bond when $R_{N-O} \leq 3.2 \text{Å}$, $\theta_{N-OH} \leq 35^\circ$ and $E \leq -3.2 \text{ kcal mol}^{-1}$, in the PCM polarization and $E \leq -3.7 \text{ kcal mol}^{-1}$ in the iterative polarization. Thus the distance between the nitrogen atoms of pyrimidine and water is less than the minimum of the radial distribution function, 3.2 Å, the angle formed between the nitrogen atom of pyrimidine and the OH bond of water is less than $35^\circ$ and the interaction energy is at least $-3.2 \text{ kcal mol}^{-1}$ (PCM) or $-3.7 \text{ kcal mol}^{-1}$ (iterative). The selection of the cut-off limit of the hydrogen bond is obtained from the histogram of the pairwise energy distribution, as seen in Fig. 6. These histograms clearly identify the hydrogen-bonded structures and show the energy minimum at $-3.2 \text{ kcal mol}^{-1}$ and $-3.7 \text{ kcal mol}^{-1}$, respectively. With these, we obtain an average of 2.1 and 2.4 hydrogen bonds formed between pyrimidine and water in the PCM and iterative polarizations, respectively.

The slightly larger number obtained in the iterative polarization is a consequence of the slightly larger calculated charge on the N atoms with a corresponding larger dipole moment, as shown in Table 1. It is instructive to show the location of these configurations in the diagram of the pairwise energy as a function of the solute–solvent distance. These are then given in Fig. 7. These diagrams clearly demonstrate that the geometry criterion alone is not enough to characterize hydrogen bonds because it can be seen that some solute–solvent energies are positive and, as such, cannot be attributed to a hydrogen bond. Only after using the energy criterion a clear hydrogen bond picture emerges.

The statistics of the hydrogen bonds formed between pyrimidine and water can be obtained in detail. Table 2 gives the entire statistics.

We find, for instance in the case of the PCM polarization that 14.1% of the configurations make one hydrogen bond, 63.6% of the configurations make two hydrogen bonds, 21.2% forms three hydrogen bonds and 1.1% forms four hydrogen bonds. These numbers for the case of the iterative polarization are also shown in Table 2. In the subsequent quantum mechanical calculations these hydrogen-bonded structures will

**Fig. 5** Radial distribution functions between the nitrogen atom of pyrimidine and the oxygen or hydrogen atom of water.

**Fig. 6** Histogram of the pairwise energy interaction between pyrimidine and water. Top (bottom) panel for the PCM (iterative) polarization.

**Fig. 7** Calculated pairwise energy interaction for pyrimidine and water showing the selection of hydrogen bonded structures. Top (bottom) panel for the PCM (iterative) polarization.
be analyzed separately. Thus in the calculations of the hydrogen bond contribution there are structures with one, two, three and four water molecules, obeying the corresponding statistics of Table 2. This is of particular interest in the QM calculations of the $^{15}$N shielding constant.

3.3 Absorption spectrum

The absorption spectrum of pyrimidine has been of interest for quite some time. With two lone pairs corresponding to each of the two nitrogen atoms, two symmetry adapted $n-\pi^*$ excitations can be obtained, corresponding to the lowest $^1B_1$ and $^1A_2$ transitions. Experimentally, the lowest $n-\pi^*$ is a broad, symmetry allowed, transition whereas the second is a dipole forbidden transition. The experimentally low-lying $n-\pi^*$ excitation has been reported in different solvents and in gas phase.77 The gas phase transition shows many narrow lines making a very broad band located in the region of ~33 900 cm$^{-1}$. Baba et al.75,76 reported the $n-\pi^*$ transition in isooctane at 34 250 cm$^{-1}$ and in water at 36 900 cm$^{-1}$. This would correspond to an isooctane–water blue shift of 2650 cm$^{-1}$. Because of the small polarity of isooctane the shift from the gas phase should be only slightly larger, close to 2700 cm$^{-1}$. Indeed the adopted suggestion has been a blue shift of 2700 ± 300 cm$^{-1}$ for the $n-\pi^*$ transition of pyrimidine in water compared to gas phase. The description of this solvent shift has been of great theoretical interest and has served also as a useful probe for testing new theoretical developments.10,64,66,69,78–80 We will analyse in detail the role of water in the solvent shift of the excitation transition. We first discuss the calculated excitation energies and these are shown in Table 3.

The lowest $n-\pi^*$ transition is calculated at 34 315 cm$^{-1}$ in the isolated case (corresponding to the gas phase and also referred to as vacuum). Comparison with the gas phase spectrum77 is more difficult because of the existence of several narrow lines in this region. The second $n-\pi^*$ transition is calculated at 37 095 cm$^{-1}$ and although dipole forbidden it has been reported experimentally77 in the gas phase in the broad range of 34 700–39 500 cm$^{-1}$. It has been argued10 that in aqueous environment there might be a near-degeneracy of these two low-lying $n-\pi^*$ transitions. For this reason we also consider in Table 3 the average value of these two transitions. But we note that in vacuum the two transitions are well separated by 2800 cm$^{-1}$, corresponding to 22 nm. We now discuss the excitation energies in water. First, we use the continuum PCM model and obtain the excitation at 35 975 cm$^{-1}$, compared with the experimental value of 36 900 cm$^{-1}$. This is in agreement with a previous result78 of 35 730 cm$^{-1}$ using the TDDFT B3LYP/6-31 + G(d) and thus indicates that the continuum PCM model with any additional consideration will give a part (~65%) of the total solvent shift. In fact previous studies10,64,66,78,80 have found the necessity of including ad hoc a few explicit water molecules, mostly to mimic the hydrogen bond interaction with the pyrimidine. Indeed, including two explicit molecules Mennucci78 obtained an improved result of 37 180 cm$^{-1}$. We consider next the explicit inclusion of the solute–solvent hydrogen bonded water molecules. The average value for the lowest $n-\pi^*$ transition is calculated at 35 050 and 34 500 cm$^{-1}$, in the two corresponding polarizations, indicating a correct trend but small blue shift. In turn, using only the discrete model gives a value that overshoots experiment (Table 3). Combining these two, by using hydrogen-bonded configurations embedded in the electrostatic field of the remaining gives a partial cancellation and a better estimate.

### Table 2
Statistics (in %) of the hydrogen bonds formed between pyrimidine and water using two (PCM and iterative) different electronic solute polarizations

<table>
<thead>
<tr>
<th>Number of HB</th>
<th>PCM</th>
<th>Iterative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14.1</td>
<td>9.1</td>
</tr>
<tr>
<td>2</td>
<td>63.6</td>
<td>43.4</td>
</tr>
<tr>
<td>3</td>
<td>21.2</td>
<td>41.4</td>
</tr>
<tr>
<td>4</td>
<td>1.1</td>
<td>6.1</td>
</tr>
<tr>
<td>Average</td>
<td>2.1</td>
<td>2.4</td>
</tr>
</tbody>
</table>

### Table 3
Results obtained for the two lowest $n-\pi^*$ excitation (in cm$^{-1}$) of pyrimidine using 100 statistically uncorrelated configurations extracted from MC simulation. Excitations obtained with the TD B3LYP/6-31 + G(d) model. Uncertainties shown are the corresponding statistical error. Also shown are the gas–water spectral shifts.

<table>
<thead>
<tr>
<th>Model</th>
<th>$E_1$</th>
<th>$\Delta E_1$</th>
<th>$E_2$</th>
<th>$E_{\text{in}}$</th>
<th>$\Delta E_{\text{in}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAS</td>
<td>34 315</td>
<td>—</td>
<td>37 095</td>
<td>35 705</td>
<td>—</td>
</tr>
<tr>
<td>PCM</td>
<td>35 975</td>
<td>1660</td>
<td>38 812</td>
<td>37 394</td>
<td>1689</td>
</tr>
<tr>
<td>PCM-QM/MM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASEC</td>
<td>37 685</td>
<td>3370</td>
<td>40 580</td>
<td>39 130</td>
<td>3425</td>
</tr>
<tr>
<td>HB</td>
<td>35 050</td>
<td>735</td>
<td>37 195</td>
<td>36 120</td>
<td>415</td>
</tr>
<tr>
<td>HB + PC</td>
<td>36 780</td>
<td>2305</td>
<td>39 535</td>
<td>38 160</td>
<td>2455</td>
</tr>
<tr>
<td>9H$_2$O + PC</td>
<td>37 105</td>
<td>2790</td>
<td>39 945</td>
<td>38 525</td>
<td>2820</td>
</tr>
<tr>
<td>28H$_2$O + PC</td>
<td>36 920</td>
<td>2605</td>
<td>39 415</td>
<td>38 170</td>
<td>2465</td>
</tr>
<tr>
<td>I-QM/MM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASEC</td>
<td>38 230</td>
<td>3915</td>
<td>41165</td>
<td>39 695</td>
<td>3990</td>
</tr>
<tr>
<td>HB</td>
<td>34 500</td>
<td>185</td>
<td>36 165</td>
<td>35 330</td>
<td>−375</td>
</tr>
<tr>
<td>HB + PC</td>
<td>36 680</td>
<td>2365</td>
<td>39 375</td>
<td>38 030</td>
<td>2325</td>
</tr>
<tr>
<td>9H$_2$O + PC</td>
<td>37 280</td>
<td>2965</td>
<td>39 925</td>
<td>38 600</td>
<td>2895</td>
</tr>
<tr>
<td>28H$_2$O + PC</td>
<td>36 955</td>
<td>2640</td>
<td>39 145</td>
<td>38 050</td>
<td>2345</td>
</tr>
</tbody>
</table>

* $\Delta E_1 = E_1 - E_{\text{gas}}$, $E_{\text{in}} = (E_1 + E_2)/2$, $\Delta E_{\text{in}} = E_{\text{in}} - E_{\text{gas}}$. 

14028 | Phys. Chem. Chem. Phys., 2010, 12, 14023–14033
The average result is within 150 cm$^{-1}$ and 250 cm$^{-1}$, compared to experiment, in the PCM and iterative polarizations respectively. This is in line with an early observation that hydrogen bonded configurations give the most important contribution but it should be supplemented with the long range effects of the outer solvent molecules. Additional improvement can be obtained by considering further the inclusion of explicit water molecules. Using 9 explicit molecules in the electrostatic embedding of 491 water molecules treated as simple charges give results close to 37 000 cm$^{-1}$ in excellent agreement with the experimental value of 36 900 cm$^{-1}$. The best result is obtained with the consideration of all solvent water molecules that comprise the first hydration shell. Hence, using similarly 28 explicit water molecules embedded in the electrostatic field of the remaining give results in sharp agreement with experiment. Comparison of these two results shown in Table 3 suggests that they are essentially converged with respect to the number of explicit solvent molecules used. Table 3 also shows that the two polarization methods are equivalent leading to the average excitation energy of 36 920 ± 90 cm$^{-1}$, using the PCM polarization and 36 955 ± 170 cm$^{-1}$ with the iterative polarization.

The numerical accuracy obtained here is excellent but, of course, we should be aware that these numbers may slightly change for different force fields or quantum mechanical models. However, some interesting conclusions can be firmly drawn. First, we note the importance of explicitly including solvent molecules. Using only the discrete model gives excitation energies that are too large and, consequently, solvent shifts that are overestimated. Also, in the same direction, including only the hydrogen bonded water as explicit solvent molecules underestimates the results, that are improved after consideration of the electrostatic embedding. The location of the hydrogen bonded water molecules should reflect the thermodynamic condition (temperature and pressure) and geometry optimised solute–solvent structures artificially overestimate the interaction. In addition, we note that the second n–π* transition also suffers a blue shift, estimated here to be in the range of 2000–2300 cm$^{-1}$. Thus in water, these two n–π* bands are not found here to be near degenerate but instead they are still separated by 2000–2350 cm$^{-1}$.

Table 4 summarizes and compares the present solvent shift of the lowest n–π* with the previous estimates. In general, different theoretical models were able to give satisfactory to very good results for the spectral shift. We note that a previous calculation without solute polarization gave an underestimated result of 1330 cm$^{-1}$, a result improved by extrapolation to the outer solvation shells. The use of the solute polarization and the electrostatic embedding improves the convergence with respect to the number of explicit solvent molecules used. We also note that the continuum model without additional considerations can give only a semi-quantitative description of the shift. However the solute polarization obtained in the PCM model seems to be reliable.

Comparison of the results obtained with the PCM and the iterative polarizations has shown that they give essentially the same result, converging to the same excitation energy of 36 900 cm$^{-1}$ corresponding to the same spectral shift of ca. 2600 cm$^{-1}$.

### 3.4 Solvent effect on the 15N magnetic shielding

After obtaining an excellent description of the one-photon optical properties of pyrimidine in water we turn to the more difficult analysis of the nuclear magnetic shielding constants. Isotropic magnetic shielding for nitrogen atoms 15N has seen an increasing interest. In this direction pyrimidine has been the focus of both experimental and theoretical studies. NMR shielding constants are very sensitive to the solute structure and to the chemical environment. Therefore they provide an interesting system for testing new theoretical solvation models. Pyrimidine in water is subjected to hydrogen bond interactions that are known to have a great influence on the isotropic magnetic shielding constants of 15N. Diazines have been the subject of a systematic experimental investigation by Witanowski et al. The gas–water shift in pyrimidine has not been considered but the cyclohexane–water change has been determined to be 16.8 ppm. Theoretically, this solvent effect was addressed in several previous studies, notably made by Mennucci and co-workers. An early investigation using PCM within the B3LYP/6-31+G(d,p) in the GIAO procedure obtained the value of 6.35 ppm for the cyclohexane–water and 10.53 ppm for the gas–water shift. In spite of the calculated low value a good correlation between theory and experiment has been obtained when considering other solvents. Note that these calculations suggest a non negligible

<table>
<thead>
<tr>
<th>Reference</th>
<th>Method/procedure</th>
<th>Spectral shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karelson and Zerner</td>
<td>SCRF, ZINDO CIS + 2 water</td>
<td>2600–3900</td>
</tr>
<tr>
<td>Gao and Byun</td>
<td>QM/MM AM1 CIS</td>
<td>2275</td>
</tr>
<tr>
<td>Zeng et al</td>
<td>MC/QM potential changes</td>
<td>1690–2450</td>
</tr>
<tr>
<td>Serrano-Andrés et al.</td>
<td>SCRF CASPT2/ANO + 2 water</td>
<td>990</td>
</tr>
<tr>
<td>Almeida et al.</td>
<td>MC/QM ZINDO CIS + 213 H2O unpolarized</td>
<td>1330</td>
</tr>
<tr>
<td>Mennucci</td>
<td>PCM TDDFT B3LYP/6-31 + G(d,p) + 1 water</td>
<td>1695</td>
</tr>
<tr>
<td>Kongsted and Mennucci</td>
<td>MD + TD DFT B3LYP/6-31 + G(d,p) + 4 water</td>
<td>1610</td>
</tr>
<tr>
<td>Present results</td>
<td>Iterative Polarization MC/QM TDDFT B3LYP/6-31 + G(d) + 28 H2O + 472 PC</td>
<td>2640</td>
</tr>
<tr>
<td>Experiments</td>
<td>2650$^b$</td>
<td>2700 ± 300$^b$</td>
</tr>
</tbody>
</table>

gas–cyclohexane shift of 4.18 ppm. The explicit use of some solvent molecules has the advantage of introducing other non-electrostatic effects not accounted by a simple PCM procedure. Improved results\textsuperscript{66,78} have been obtained when considering a few water molecules even in energy-optimized structures. The present study will include all hydrogen bonded water molecules derived from the detailed statistical analysis made in Section 3.2. Thus we not only include the necessary effects of the hydrogen bonds but in addition these are included in the proper thermodynamic condition. As an extension, further explicit solvent molecules are included and long range effects are considered in the electrostatic embedding. The results are presented in Table 5 and have considered the two electronic solute polarizations. Early theoretical studies\textsuperscript{66} have shown that PCM have a tendency to underestimate the solvent shift of diazines. For instance, the previous PCM result\textsuperscript{66} for the gas–water chemical shift of pyrimidine is 12.2 ppm. Most of this has been attributed to the inadequacy of the treatment of the effects of the hydrogen bonds between the pyrimidine and water. First, we note that the use of the solute and the hydrogen-bonded water molecules seem to give a very good agreement with the experimental results. For instance, by using the PCM polarization in the MC simulation the result obtained for the gas–water chemical shift is 17.4 ppm in very good agreement with the cyclohexane–water experimental value of 16.8 ppm. A next step for improvement is the consideration of the electrostatic embedding provided by the remaining solvent molecules. Thus we consider two additional procedures. In the first, these hydrogen bonded solvent molecules are embedded in the electrostatic field of the remaining solvent molecules. This increases the value of the chemical shift and gives results (24.5 ppm in the case of the PCM polarization) that seem to overestimate the experimental value. As the ASEC results show the sole electrostatic contribution is very large and this is reflected in the HB + PC results. A more realistic model is provided by the explicit use of the hydrogen bonded solvent molecules plus some additional explicit molecules. This however, shows little, if any, improvement. In fact these 9H2O + PC results indicate that the hydrogen bonded solvent molecules are the most important contributors to the chemical shift and additional inclusion of explicit molecules is unnecessary. We partially conclude that the explicit use of the hydrogen bonded water molecules is crucial for a good description and that the PCM polarization leads to a better result for the solvent effect on this particular property. Our best result points to a solvent effect that gives a gas–water 15N NMR chemical shift of 24.4 ppm. As stated before, the experimental value for the cyclohexane–water shift is 16.8 ppm. This leaves open an analysis of the gas–cyclohexane shift, necessary for a better comparison between theory and experiment. Fig. 8 shows a plot of the experimental results of ref. 82 for the 15N NMR chemical shift of pyrimidine as a function of the Reichardt\textsuperscript{84} normalized solvent polarity, for twelve different solvents. We may note that extrapolation to the gas phase (ETN = −0.111)\textsuperscript{84} is possible, although subjected to some uncertainty. This extrapolation suggests a gas–water chemical shift value of 18.5 ± 2 ppm. For comparison, adding a previous theoretical estimate of the gas–cyclohexane shift of 4.18 ppm\textsuperscript{78} to the experimental cyclohexane–water shift would place the gas–water chemical shift around 21 ppm.

All considered, we assume the experimental gas–water chemical shift of 15N in pyrimidine to be in the range of 17–21 ppm. This would make a good agreement with our theoretical results (9H2O + PC), in special, using the PCM polarization. In the iterative polarization there is some overestimation of ca. 7–8 ppm. Given the difficulties of the precise theoretical determination of the solvent effects on the 15N shielding of pyrimidine, we may conclude that these are very good results. The PCM polarization, used in conjunction with the sequentially combined MC and QM calculations, is seen to be a successful alternative.

We now briefly discuss some possible sources of inaccuracies in the calculations of the 15N magnetic shielding of pyrimidine. As the geometry used here has been optimized in vacuum we first discuss how this affects the quality of our results. Geometry optimization in water environment could be made using either PCM or SCRF but because of the relative stiffness of pyrimidine this is expected to give only minor differences in geometry with mild changes on the 15N chemical shift. This will not be the case, however, if specific hydrogen-bonded solvent molecules are included in the optimization. In this case, smaller N–HOH distances are obtained and the distortion around the nitrogen site of pyrimidine would be more

Table 5  Calculated gas–water chemical shift (in ppm) of the 15N isotropic magnetic shielding of pyrimidine using different polarizations and the B3LYP/GIAO/aug-pcS2(N)/pcS2(C,O,H) model. For reference, the calculated 15N isotropic magnetic shielding value in the gas phase is 91.7 ppm and the continuum PCM shift is calculated as 15.4 ppm.

<table>
<thead>
<tr>
<th>Polarization</th>
<th>ASEC</th>
<th>HB</th>
<th>HB + PC</th>
<th>9H2O + PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCM</td>
<td>25.8</td>
<td>17.4 ± 1.0</td>
<td>24.5 ± 0.8</td>
<td>24.4 ± 0.8</td>
</tr>
<tr>
<td>Iterative</td>
<td>30.4</td>
<td>16.4 ± 0.7</td>
<td>27.8 ± 0.7</td>
<td>28.5 ± 0.8</td>
</tr>
</tbody>
</table>

Experimental value\textsuperscript{82} for the cyclohexane–water shift is 16.8 ppm. Calculated\textsuperscript{78} gas–cyclohexane shift using IEF PCM is 4.2 ppm. Suggested experimental gas–water chemical shift is 17–21 ppm. See text.

Fig. 8 Dependence of the experimental 15N chemical shift of pyrimidine with the Reichardt normalized polarity\textsuperscript{84} for twelve different solvents. Numerical values from the experiments of Witanowski et al.\textsuperscript{82} with different solvents: 1 = cyclohexane, 2 = CCl4, 3 = benzene, 4 = EtO, 5 = dioxane, 6 = CHCl3, 7 = CH2Cl2, 8 = acetone, 9 = DMSO, 10 = EtOH, 11 = MeOH, 12 = H2O.
appreciable. However, based on previous studies this will likely increase\textsuperscript{66} the chemical shift, the opposite to what is needed to achieve even better agreement with experiment.

A clear source of uncertainty also comes from the neglect of zero-point vibrational correction (ZPVC). An early study\textsuperscript{85} has indeed shown that ZPVC has a sizable contribution. But has obtained equivalent values in gas and in water, so there was a mutual cancellation leading only to a small, residual, contribution.

We then consider the theoretical model used, B3LYP/aug-\textit{pcS}2(N)/\textit{pcS}2(C,O,H). The B3LYP has been argued to have a tendency to overestimate chemical shifts and could be one of the main sources of error here. But theoretical examples, for the \textsuperscript{15}N NMR chemical shift of NH\textsubscript{3} in gas\textsuperscript{36} and in homogeneous liquid ammonia\textsuperscript{87} have shown a good performance of the B3LYP in conjunction with the specially designed aug-\textit{pcS}-\textit{X} basis sets. In turn, we cannot attribute large deficiencies in the compromising basis set because our test calculations have shown that the difference between the results using the aug-\textit{pcS}2(N)/\textit{pcS}2(C,O,H) basis set is less than 1 ppm if the full aug-\textit{pcS}-3 is used for all atoms. Finally, the absence of the solvent polarization by the solute may have some small effect on the \textsuperscript{15}N shielding. This is more difficult to analyze but recent work\textsuperscript{88} has shown it to have relatively small importance in the more sensitive case of spin–spin coupling constants.

As a last consideration we should observe that the PCM polarization implemented in the MM part (a Monte Carlo simulation, here) seems to generate an equivalent or even better description of the solute-solvent hydrogen bonds, judging from the better agreement with experiment for the \textsuperscript{15}N NMR chemical shift of pyrimidine (Table 5). Whether this is a more general conclusion it is difficult to ascertain at this stage. But clearly the PCM method provides a good and economical average solute polarization.

Before concluding, we show in Fig. 9 the statistical convergence of the average n−π\textsuperscript{*} transition and the isotropic \textsuperscript{15}N chemical shielding of pyrimidine in aqueous environment.

### 4. Conclusions

The solvent effects on the low-lying n−π\textsuperscript{*} absorption transitions and the \textsuperscript{15}N isotropic chemical shielding of pyrimidine in water have been calculated using the sequential Monte Carlo/Quantum mechanical (S-QM/MM) model. As the classical and quantum mechanical parts are uncoupled in the sequential procedure a special consideration of the solute polarization is needed. This is made using two different approaches. In the first, we use an iterative method that electrostatically equilibrates the solute to the solvent. Second, the possibility of obtaining the average solute polarization from the polarizable continuum model has also been used as an alternative. The two different approaches—PCM and QM/MM methods—are then simultaneously used and reconciled. We use PCM to obtain the average electronic solute polarization by the solvent and include this in the MM part of the S-QM/MM methodology. We analyze the solute polarization by monitoring the in-water dipole moment. The calculated values are 3.89 D (PCM polarization) and 4.08 D (iterative polarization), corresponding to an increase of 61\% and 69\% compared to the calculated gas phase value of 2.42 D. The two polarizations are also used to analyze the structural aspects and the solute–solvent hydrogen bonds. The PCM polarization gives an average number of 2.1 solute–solvent hydrogen bonds, compared to the iterative procedure that gives the slightly larger average number of 2.4. Finally we consider the electronic absorption spectrum and the nuclear magnetic shielding, with separate analysis of the continuum, discrete and explicit solvation models. In this latter we also separate the effects of the solute–solvent hydrogen bonds from the bulk solvent water molecules. The best results obtained here use the solute pyrimidine surrounded by explicit water molecules embedded in the electrostatic field of the remaining.

For the low lying n−π\textsuperscript{*} absorption transition in water we use 28 explicit water molecules and obtain the excitation energy of 36 900 ± 100 (PCM polarization) and 36 950 ± 100 cm\textsuperscript{-1} (iterative polarization), in excellent agreement with the band maximum observed at 36 900 cm\textsuperscript{-1}. For the nuclear shielding \textsuperscript{15}N we use 9 explicit water molecules to obtain the corresponding gas–water chemical shifts values of 24.4 ± 0.8 and 28.5 ± 0.8 ppm, compared with the inferred experimental value of 17–21 ppm. A careful consideration has been made of the possible inaccuracies of the calculated \textsuperscript{15}N chemical shielding. To proceed in this direction, additional experimental data using solvents of low polarity would be useful. Analysis of the hydrogen bond contribution to the total solvent effect is made considering the corresponding statistics and indicates that these are by far the most important contribution to the
NMR chemical shift, with only residual contributions coming from the other solvent molecules. Finally, all theoretical results obtained here are verified to be statistically converged.

Considering the simplicity of the PCM over the iterative polarization the agreement obtained comparing these two procedures is an important aspect leading to computational savings that suggest the possibility of using this PCM-MM/QM model to treat larger solute molecules in solvent environment. This is underway for optical spectrum of biomolecules in water and presenting similarly good agreement.

Acknowledgements

This work has been partially supported by the Institute of Complex Fluids (INCT-FCx). We also acknowledge continuous support from CNPq, CAPES, FAPEAL and FAPESP (Brazil).

Notes and references

83 B. Mennucci, ch. 1 in ref. 22, pp. 1–21.