An efficient quantum mechanical/molecular mechanics Monte Carlo simulation of liquid water

Willian R. Rocha [a,c], Kaline Coutinho [b], Wagner B. de Almeida [a,c], Sylvio Canuto [d,*]

[a] *Núcleo de Estudos em Química Computacional (NEQC), Departamento de Química, ICE, UFJF, Juiz de Fora, MG 36036-330, Brazil*
[b] *Centro de Ciências Exatas e Tecnológicas (CCET), Universidade de Mogi das Cruzes, CP 411, Mogi das Cruzes, SP 08701-970, Brazil*
[c] *Laboratório de Química Computacional e Modelagem Molecular (LQC-MM), Departamento de Química, ICEx, UFMG, Belo Horizonte, MG 31270-901, Brazil*
[d] *Instituto de Física Universidade de São Paulo, CP 66318, São Paulo, SP 05315-970, Brazil*

Abstract

An efficient implementation of a quantum mechanical/molecular mechanics (QM/MM) Monte Carlo (MC) simulation of liquid water is presented. The auto-correlation function of the energy is used to reduce the number of QM calculations during the simulation. The effects of the liquid structure on the dipole moment of the quantum water molecule is considered. Using only 80 QM calculations at the Hartree–Fock/6-31++G(d,p) level but on statistically nearly uncorrelated structures, we find an average dipole moment for the liquid water of 2.73 ± 0.13 D. This corresponds to an induced dipole moment of 0.44 ± 0.13 D on going from the gas phase to the liquid phase, in excellent agreement with previous, more extensive computational procedures. © 2001 Elsevier Science B.V. All rights reserved.

1. Introduction

Computer simulation has allowed the understanding of the intermolecular interactions and the behavior of organic molecules and biological macromolecules in solution. Statistical mechanics techniques such as Monte Carlo (MC) [1] and molecular dynamics (MD) [1] have been widely used to study the structure and thermodynamical properties of pure liquids, solvation phenomena and the solvent effects on structural, thermodynamical and spectroscopic properties of molecules

in solution. The most commonly used potentials in such methods are of the molecular mechanics (MM) type which, by their nature, do not allow the study of processes in which a quantum mechanical (QM) treatment is needed such as, for example, excited states, bond breaking and bond making.

In the last few years, the hybrid quantum mechanical/molecular mechanics (QM/MM) approach has been emerging as a powerful method for studying processes in condensed phase [2–13]. In such a procedure, the solute molecule (or the portion of the system that needs a QM description) is treated quantum mechanically, while the remainder of the system (solvent molecules) is represented by classical force fields. These two

* Corresponding author. Fax: +55-11-38186831.
E-mail address: canuto@if.usp.br (S. Canuto).

subsystems communicate through electrostatic and van der Waals interactions terms. In this approach, electronic structure calculations are carried out during the simulation and energy as well as energy derivatives of the quantum subsystem are obtained. This QM/MM approach has been implemented in both MC [8] and MD [3] methods.

In normal MC sampling procedures, the number of steps required to obtain the averages is about 10^6. Then, the number of QM calculations required in a hybrid QM/MM simulation is too high, since in every step an energy evaluation of the system is needed when an electronic structure calculation is performed. So, a great computational effort is necessary to carry out this kind of simulation. Aiming at reducing the computational demand, previous implementations of this method have utilized more inexpensive electronic structure calculations on the quantum part as, for example, valence-bond [2,14] and semi-empirical molecular orbital methods [3–5]. However, for some classes of compounds, for example transition metal compounds, in which we are interested [15], these approximate molecular orbital methods are not appropriate due to the lack of parameters. To extend the applicability of the QM/MM hybrid simulations for more general systems, a consistent scheme to reduce the number of QM calculations during the simulation is relevant. A step in this direction has recently been given by Truong et al. [12,13] in their ab initio perturbative approach implementation and Tuñón et al. [16] in their DFT implementation of the QM/MM method.

In this work we present an implementation of the QM/MM method in an MC NVT simulation of a quantum water molecule in a box of classical water molecules. The auto-correlation function of the energy is introduced to reduce the number of ab initio QM calculations during the simulation without compromising the numerical accuracy of the average properties. As a test example we calculate the dipole moment of liquid water. This seems to be an appropriate test as several previous procedures have been employed before [4,16–20] with varying computational complexities.

2. Computational details and methodology

In the QM/MM approach, the entire system under investigation is partitioned into regions, which are treated differently in the course of the simulation. The molecular system is divided into a QM region, an MM region, and a boundary region. Detailed descriptions of the development and applications of the QM/MM approach have been given in recent reviews [8,11].

2.1. MC simulation

The MC simulation is performed using standard procedures for the Metropolis sampling technique in the canonical NVT ensemble, where the number of molecules N, the volume V and the temperature T are fixed. As usual, periodic boundary conditions in a cubic box [1] are used. The system consisted of 216 TIP3P [21] water molecules in a cubic cell with a side length of 18.627 Å. The volume of the cube was determined by the density of the liquid water ($\rho = 0.996$ g/cm^3) and the temperature was fixed at 298 K. Periodic boundary conditions and a cutoff distance of 9.314 Å have been applied for all classical–classical and classical–quantum interactions and the long-range corrections of the Lennard–Jones potential were calculated beyond this cutoff distance. Metropolis sampling was supplemented by preferential sampling, using the Owicki–Scheraga procedure [22], with a solvation radius of 7.451 Å, which is 80% of the cutoff radius, producing a probability factor of 0.34 to be used in the preferential sampling. The simulation consisted of a thermalization stage of 1.08×10^6 MC steps, followed by an additional 10.7×10^6 MC steps, in which the QM/MM was turned on. A new configuration is generated after 215 MC steps, i.e. after all solvent molecules attempt to translate and rotate around a randomly chosen axis. This generates 48 000 MC configurations. The maximum displacement of the molecules is self-adjusted after 50 configurations, to give an acceptance ratio around 50%.

The geometry and the Lennard-Jones parameters of the water molecules were those of the

TIP3P water monomer. The quantum water molecule was treated at the Hartree–Fock (HF) level of theory, using the 6-31++G(d,p) basis set, which includes polarization and diffuse functions on the oxygen and hydrogen atoms.

2.2. Statistical correlation

As the appropriate Boltzmann weights are included in the Metropolis MC sampling technique [1], the average value of a property such as the energy is given as a simple average over a chain of size L of instantaneous values:

$$\langle E \rangle_L = \frac{1}{L} \sum_i^L E_i. \tag{1}$$

Since L is large enough, the average value of the property converges and the error caused by the finite size of the simulation is given by

$$\sigma(\langle E \rangle_L) = \sqrt{\frac{\langle E^2 \rangle_L - \langle E \rangle_L^2}{L-1}}, \tag{2}$$

if the instantaneous values, E_i, are statistically independent or uncorrelated. Of course, successive configurations are usually not independent, because each molecule was attempted to move only once. Therefore, it is necessary to calculate the interval of uncorrelated configurations to have a statistical efficiency in the calculation of average properties and to calculate the statistical error properly [23].

There are two independent methods of calculating the interval of uncorrelated configurations. One method uses the auto-correlation function of the energy to obtain the correlation step τ [24] and the other uses the statistical inefficiency s [25]. The auto-correlation function, $C(n) = \langle \delta E_i \delta E_{i+n} \rangle / \langle \delta E^2 \rangle$, gives information on the statistical correlation of successive configurations. It is known that for Markovian process, $C(n)$ has an exponential decay [24] with $[C(n) = \exp(-n/\tau)]$ or $[C(n) = \sum_i c_i \exp(-n/\tau_i)]$, and the interval between uncorrelated configurations, the so-called correlation step τ, is calculated as an integration from zero to infinity of the $C(n)$. Of course, for an exponential decay, truly uncorrelated configurations

will require an infinite separation, i.e. $C(n) = 0$, only in infinity. In practical terms, configurations separated by 2τ, or larger, are considered uncorrelated.

The statistical inefficiency, $s = \lim_{n \to \infty} (n \langle \delta E^2 \rangle_n / \langle \delta E^2 \rangle)$, is based on the mean-square deviations of averages taken over blocks of a chain of configurations and the limiting value s signifies that the block size n has become so large that there is no correlation from block to block. Therefore in practice, the configurations separated by interval larger than s are uncorrelated.

Even though these two methods of calculating the interval of uncorrelated configurations are independent, they can be related by $s = 2\tau$ [1]. In practice, for a Markovian process, an interval of successive configurations greater than 2τ gives uncorrelated configurations.

Both methods were already used to improve the calculation of some thermodynamic properties such as pressure, chemical potential [26] and equilibrium constant [27]. In early works [28–30], we have used them to reduce the number of configurations used in the calculation of the solvatochromic shifts of some solutes in solvents.

Here, we use the auto-correlation function of the energy to reduce the number of electronic structure calculations during the simulation using only the uncorrelated configurations. In Fig. 1, we show the calculated auto-correlation function and
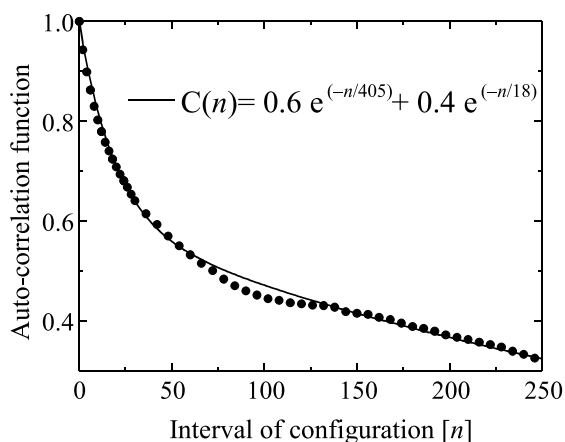

Fig. 1. Calculated auto-correlation function of the energy for water in liquid water and best exponential fit.

it best fits as a sum of two exponential decay, $C(n) = 0.6 \exp(-n/405) + 0.4 \exp(-n/18)$. Integrating the best fit, we obtain $\tau \sim 280$ and $s \cong 560$. It means that from the 48 000 MC configurations generated during the simulation, we selected 80 uncorrelated configurations separated by an interval of 560 configurations. In [28], we show that the final average of the solvatochromic shift of benzene in water is not compromised with the reduction in the number of configurations considered. Using 20 000 correlated MC configurations or 25 uncorrelated MC configurations gave the same result. Here we will also show a numerical demonstration of the convergence of the calculated dipole moment.

The correlation function of the energy is therefore an important parameter in selecting and reducing the number of configurations to be used in QM calculations. We have recently obtained the correlation step for some common solvents [30]. The HF ab initio calculations were carried out in intervals of 560 MC configurations. This yielded 80 QM calculations on nearly uncorrelated configurations in which the dipole moment was averaged out. Each configuration is composed of 215 TIP3P classical water molecules and the reference quantum mechanical water.

Our implementation involved a modification of the GAMESS QM code [31] which was interfaced with our DICE [32] MC statistical mechanics program.

3. Results and discussion

3.1. Dipole moment of water

Our interest here is not to obtain an improved dipole moment of liquid water. Instead, we are interested in examining the feasibility of reducing the total number of QM calculations without affecting the average property of interest. In this direction, the dipole moment of water in aqueous solution is appropriate as there are several theoretical results to compare with [4,16–20]. The induced dipole moment, i.e. the difference between the dipole moment in the liquid and in the gas, is of interest here. Perusal of the different theoretical models used before [4,16–20] have suggested that this induced dipole moment varies between 0.35 and 0.65 D with the experimental value [33,34] being estimated as 0.75 D. Results using HF or DFT with double-zeta basis [17,20] indicate that at this level the induced dipole moment is 0.45 D. This is the estimated value we should obtain using our HF/MM with the 6-31++G(d,p) basis.

The calculated and experimental dipole moment for the water molecule in gas phase and in solution are given in Table 1. The calculated gas phase dipole moment for the water monomer is usually overestimated compared with the experimental value. The use of a triple-zeta basis with field-induced polarization functions (TZPV+) in the density function theory using the local density

Table 1
Calculated and experimental dipole moment (in Debye) for the water molecule in the gas phase and in solution[a]

Method	Reference	Potential	Gas phase	Liquid phase	$\langle\Delta\mu\rangle$
Semi-empirical AM1/MM	4	SPC and TIP3P	1.86	2.15	0.29 ± 0.05
DFT–LDA–DZVP/MM	17	SPC and TIP3P	2.28	2.71	0.43 ± 0.04
DFT–LDA–TZVP+/MM	17	SPC and TIP3P	1.88	2.54	0.66 ± 0.05
DFT–Perdew–DZVP/MM	17	SPC and TIP3P	2.17	2.62	0.45 ± 0.04
DFT–Perdew–TZVP+/MM	17	SPC and TIP3P	1.81	2.47	0.66 ± 0.05
DFT–VWN–Becke–Perdew/MM	16	TIP3P	2.28	2.89	0.61 ± 0.05
Car–Parrinello MD	19	Ab initio	–	2.95	–
Polarizable water potential	18	SPC	1.85	2.81	1.00 ± 0.25
HF/6-311G(d,p) QM/MM	20	Dimer optimized	2.14	2.49	0.35 ± 0.13
HF/6-311G(d,p) QM/MM	20	TIP3P	2.14	2.79	0.65 ± 0.16
HF/6-31++G(d,p) QM/MM	Present	TIP3P	2.29	2.73	0.44 ± 0.13
Experiment	33, 34		1.85	2.60	0.75

[a] $\Delta\mu$ is the difference between the dipole moment in gas and in water. See text.

approximation (LDA) or Perdew non-local potential [17] gives a good value for the gas phase dipole moment. When the quantum water molecule is placed into the classical liquid, the electronic density is substantially polarized, which will affect the dipole moment. As can be seen in Table 1, our average dipole moment obtained at the end of the simulation is $\langle \mu \rangle = 2.73$ D, which gives an average induced dipole moment of 0.44 D on going from the gas phase to the liquid phase.

The precise value of the water dipole moment in the liquid phase is still an open question. The usual value taken for parameterization of potential energy functions is $\langle \mu \rangle = 2.6$ D [33]. Using an isotropic polarizable water model, Bernardo et al. [18], have obtained a value of $\langle \mu \rangle = 2.81$ D for the liquid water with the large value for the induced dipole moment of 1.00 D. Tuñón et al. [16] have obtained an average value of $\langle \mu \rangle = 2.89$ D, and an induced dipole moment of 0.61 D, in their DFT/MM hybrid simulation of liquid water. They used the Vosko–Wilk–Nusair functional with Becke–Perdew corrections [16]. Silvestrelli and Parrinelo [19] obtained directly the liquid dipole moment with an average value of $\langle \mu \rangle = 2.95$ D using ab initio molecular dynamics calculation on a cluster of 64 water molecules. For such a large computational effort, they used an average of only 12 molecular dynamics configurations. Recently, Tu and Laaksonen [20] obtained a very good induced dipole moment value of 0.65 D in their QM/MM molecular dynamics study of liquid water, treating the quantum molecule at the HF/6-311G(d,p) level and using the TIP3P potential to describe the interactions within the MM region. Our HF/6-31++G(d,p) QM/MM MC simulation gave an average value of the water dipole moment which is essentially the same obtained using the similar HF/6-311G(d,p) QM/MM molecular dynamics [20], in which a much greater number of QM calculations was performed during the simulation. However, [20] obtains an improved induced dipole moment. The calculations of Wei and Salahub [17] using a basis set of the same quality as ours gave an induced dipole moment of 0.43 D within the LDA approximation and 0.45 D using the Perdew potential. This result is indeed very close to our result of 0.44 D. They [17] have used around 1000 QM

calculations. With a better basis set they obtained 0.66 D both in LDA and Perdew approximations. These facts are indicative that our approach to reduce the number of electronic structure calculations during the QM/MM simulation produces very satisfactory results, with a much lower computational cost. Fig. 2 shows the histogram of the calculated dipole moment of the quantum water molecule.

At this stage it may be convenient to analyze the dependence of our calculated average value of the dipole moment of water on the set of L values used in the calculation. As discussed before, configurations separated by 2τ may be considered uncorrelated and therefore the average value is converged and should be independent of the particular choice of the set of MC configurations selected for the QM calculations. The error, however, depends on the total number of configurations used to calculated the average, that is on the total size of the simulation. Table 2 shows the calculated average dipole moment for different sizes L of the simulation but with the same correlation step, or statistical inefficiency. The results of Table 2 clearly demonstrate that the average value has converged but the error σ decreases with increasing L. In our numerical test this amounts to a factor of nearly three for $L = 10$ and $L = 80$. Similarly, using configurations separated by 3τ, gives the value of 2.75 ± 0.15 D. As it can be noted, this is essentially the same converged value of 2.73 ± 0.13 D but it
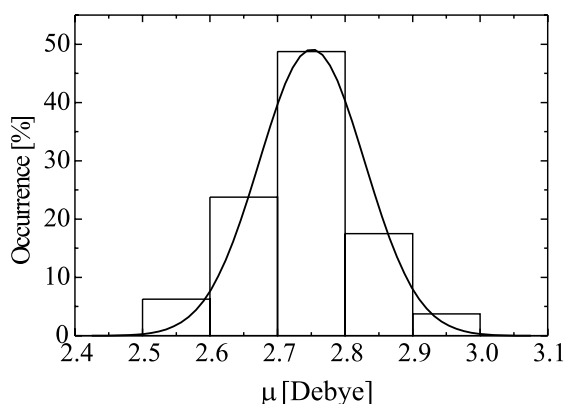

Fig. 2. Histogram and Gaussian distribution of the calculated dipole moment of water in liquid water.

Table 2
Convergence of the calculated average dipole moment of water $\langle\mu\rangle$ with the number L of instantaneous values (see Eq. (1))[a]

MC steps (In 10^6)	L	$\langle\mu\rangle \pm \sigma$
1.20	10	2.75 ± 0.36
2.41	20	2.67 ± 0.30
3.61	30	2.68 ± 0.22
4.82	40	2.71 ± 0.19
6.02	50	2.73 ± 0.16
7.22	60	2.73 ± 0.15
8.43	70	2.73 ± 0.14
9.63	80	2.73 ± 0.13

[a] All configurations used in the average are separated by $s = 2\tau$ (see text). σ is the statistical error: $\sigma(\langle\mu\rangle_L) = \sqrt{\frac{\langle\mu^2\rangle_L - \langle\mu\rangle_L^2}{L-1}}$.

slightly affects the error σ because this decreases the size L of instantaneous values used for the average.

3.2. Radial pair distribution function

The radial distribution function (RDF) reflects in a complementary way the ability of the present approach in representing the liquid structure and potential. Fig. 3 shows the computed RDF obtained for the quantum water molecule. The numerical values for the main peak positions are shown in Table 3. Fig. 3 and Table 2 also show the experimental neutron diffraction results obtained for the liquid water at 25°C [35].

Hydrogen bonding interactions are apparent from the first sharp and strong peak in the $G_{OH}(r)$ (Fig. 3a). Our calculated $G_{OH}(r)$ are in good agreement with the experimental findings in both peak positions (see Table 3) and relative intensities. Integration of the first peak gives an average number of hydrogen bonds for the quantum water molecule of 3.06, which is in reasonable agreement
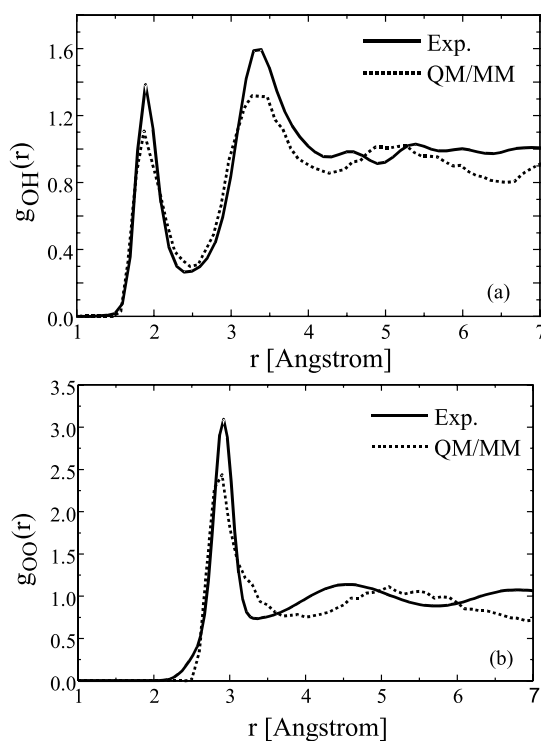

Fig. 3. Calculated and experimental [35] radial distribution functions. (a) Quantum oxygen–TIP3P hydrogen (normal line) and experimental (dotted line). (b) Quantum oxygen–TIP3P oxygen (normal line) and experimental (dotted line).

with the experimental result of 3.58 [35]. Despite the fact that the intensities of the peaks obtained in the $G_{OH}(r)$ (Fig. 3a) are slightly underestimated compared with the experimental result, the relative intensities are in good agreement with the experiment. Similar results were obtained recently by Tu and Laaksonen [20]. The previous DFT/MM study [16] has obtained a good representation of the peak positions but with inverted relative intensities.

Table 3
Experimental and calculated peak positions in the radial distribution function for the liquid water

	Experiment [35]	QM/MM HF/6-311G(d,p) Ref. [20]	QM/MM HF/6-31++G(d,p) Present results
RDF first peak			
$g_{OO}(r)$	2.88	2.66	2.85
$g_{OH}(r)$	1.85	1.74	1.85
RDF second peak			
$g_{OO}(r)$	4.50	5.00	5.00
$g_{OH}(r)$	3.25	3.15	3.25

The $G_{OO}(r)$ has the correct structure, compared with the experimental results, showing two clearly distinct peaks. As it can be seen from Table 2, the first peak occurs at 2.85 Å, which is in good accord with the experimental value of 2.88 Å. The second peak at 5.00 Å, appears at a slightly larger distance than the experimental value of 4.50 Å. There is a general agreement between our calculated QM/MM structural results for the liquid water and those obtained experimentally.

4. Conclusions

In this work we have presented an implementation of a QM/MM Monte Carlo NVT simulation of a quantum water molecule in a box of classical water molecule. The auto-correlation function of the energy was applied in this implementation of the QM/MM, to reduce the number of QM calculations during the simulation. The quantum water molecule was treated at the HF/6-31++G(d,p) level of theory, surrounded by classical water molecules described by the TIP3P potential. We have found an average dipole moment for the liquid water of 2.73 ± 0.13 D, which gave an averaged induced dipole moment of 0.44 D on going from the gas phase to the liquid phase. These results strongly suggest that the analysis of the auto-correlation function is a promising procedure to optimize the statistics of the simulation and reduce considerably the computational effort in QM/MM studies.

Acknowledgements

References

[1] M.P. Allen, D.J. Tildesley, Computer Simulation of Liquids, Oxford University Press, Oxford, 1987.
[2] A. Warshel, M. Levitt, J. Mol. Biol. 103 (1976) 227.
[3] M.J. Field, P.A. Bash, M. Karplus, J. Comp. Chem. 11 (1990) 700.
[4] J. Gao, X. Xia, Science 258 (1992) 631.
[5] J. Gao, J. Phys. Chem. 96 (1992) 537.
[6] J. Åqvist, A. Warshel, Chem. Rev. 93 (1993) 2523.
[7] D. Bakowies, W. Thiel, J. Phys. Chem. 100 (1996) 10580.
[8] J. Gao, In: K.B. Lipkowitz, B.B. Boyd (Eds.), Reviews in Computational Chemistry, VCH, New York, 1996, p. 119.
[9] R.V. Stanton, D.S. Hartsough, K.M. Merz Jr., J. Comp. Chem. 16 (1995) 113.
[10] M.A. Thompson, J. Phys. Chem. 100 (1996) 14494.
[11] J. Gao, Acc. Chem. Res. 29 (1996) 298.
[12] T.N. Truong, E.V. Stefanovich, Chem. Phys. Lett. 256 (1996) 348.
[13] T.J. Evans, T.N. Truong, J. Comp. Chem. 19 (1998) 1632.
[14] A. Warshel, Computer Modeling of Chemical Reactions in Enzymes and in Solution, Wiley, New York, 1991.
[15] W.R. Rocha, W.B. De Almeida, Organometallics 17 (1998) 1961.
[16] I. Tuñón, M.T.C. Martins-Costa, C. Millot, M.F. Ruiz-López, J.L. Rivail, J. Comp. Chem. 17 (1996) 19.
[17] B. Wei, D.R. Salahub, Chem. Phys. Lett. 224 (1994) 291.
[18] D.N. Bernardo, Y. Ding, K. Krogh-Jespersen, R.M. Levy, J. Phys. Chem. 98 (1994) 4180.
[19] P.L. Silvestrelli, M. Parrinello, Phys. Rev. Lett. 82 (1999) 3308.
[20] Y. Tu, A. Laaksonen, J. Chem. Phys. 111 (1999) 7519.
[21] W.L. Jorgensen, J. Chandrashekar, J. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926.
[22] J.C. Owicki, H.A. Scheraga, Chem. Phys. Lett. 47 (1977) 600.
[23] G. Jacucci, A. Rahman, N. Cimento D 4 (1984) 341.
[24] C. Chatfield, The Analysis of Time Series: An Introduction, Chapman and Hall, London, 1984.
[25] R. Friedberg, J.E. Cameron, J. Chem. Phys. 52 (1970) 6049.
[26] D. Fincham, N. Quirke, D.J. Tildesley, J. Chem. Phys. 84 (1986) 4535.
[27] W.C. Swope, H.C. Andersen, P.H. Berens, K.R. Wilson, J. Chem. Phys. 76 (1982) 637.
[28] K. Coutinho, M.J. Oliveira, S. Canuto, Int. J. Quant. Chem. 66 (1998) 249.
[29] S. Canuto, K. Coutinho, Int. J. Quant. Chem. 77 (2000) 192.
[30] K. Coutinho, S. Canuto, M.C. Zerner, J. Chem. Phys. 112 (2000) 9874.
[31] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S.J. Su, T.L. Windus, M. Dupuis, J.A. Montgomery, J. Comp. Chem. 14 (1993) 1347.
[32] K. Coutinho, S. Canuto, DICE: A Monte Carlo Program for Molecular Liquid Simulation, University of São Paulo, Brazil, 1997.
[33] E. Whalley, Chem. Phys. Lett. 53 (1978) 449.
[34] S.A. Clough, Y. Beers, G.P. Klein, L.S. Rothman, J. Chem. Phys. 59 (1973) 2554.
[35] A.K. Soper, F. Bruni, M.A. Ricci, J. Chem. Phys. 106 (1997) 247.