

Sampling Configurations in Monte Carlo Simulations for Quantum Mechanical Studies of Solvent Effects

KALINE COUTINHO, M. J. DE OLIVEIRA, SYLVIO CANUTO

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

Received 11 March 1997; accepted 22 August 1997

ABSTRACT: Discrete models obtained from computer simulation are in increase use to study solvent effects. The approach consists of generating supermolecular structures for quantum mechanical calculations. The properties of the solute are calculated as an ensemble average over configurations generated by the simulation. An analysis of the efficiency of the simulation shows that the number of configurations necessary for the ensemble average can be reduced drastically. As an application to solvatochromism, the calculated spectral shift of the ${}^1B_{2u}(\pi - \pi^*)$ transition of benzene in water is shown to be the same whether it is calculated with many but correlated configurations or with just a few uncorrelated configurations. © 1998 John Wiley & Sons, Inc. *Int J Quant Chem* 66: 249–253, 1998

Introduction

It has long been recognized that solvent effects would be a central topic in theoretical physical chemistry [1]. This is, of course, a simple recognition that most experimental data are obtained in solution. An approach that is gaining increased popularity is the use of Monte Carlo or molecular dynamics simulation to generate supermolecular structures of the solute surrounded by the solvent molecules that are used next in quantum mechanical (QM) calculations [2–5]. As the liquid has many

structures or configurations (geometrical arrangement of the molecules) equally possible at a certain temperature, some sort of ensemble average is necessary to describe the system in the liquid phase. However, as the simulation produces several thousands of configurations, the subsequent QM calculation is not feasible if the number of necessary configurations is not greatly reduced. Some investigators [3] have simply substituted the discrete solvent molecules by classical point charges. This procedure neglects all the effects of solvent polarization. Consideration of the solvent polarization is indeed necessary [2, 6]. Obviously, the number of solvent molecules around the solute has also to be drastically reduced but this can be judiciously done by including solvation shells ac-

Correspondence to: K. Coutinho.

Contract grant sponsors: CNPq; FAPESP.

cording to the pair radial distribution function (RDF) [5]. We seek a statistical procedure based on the concept of efficiency of the simulation to reduce the number of configurations included in the QM calculations, as the intention is to treat the solute, the solvent, and its interaction by quantum mechanics. Therefore, in this article, we suggest that instead of performing QM calculations in every configuration generated by the (Monte Carlo) simulation this can be performed only on a few uncorrelated configurations without compromising the final average result. We demonstrate that this can be accomplished with the use of a quantity that measures the statistical inefficiency. As the working example, we use the interesting case of the solvatochromic shift of the first $B_{2u}(\pi - \pi^*)$ absorption band of benzene in water.

Monte Carlo Simulation and Statistical Inefficiency

The Monte Carlo (MC) simulation [7] is carried out with N rigid molecules in a cubic box of volume V at a fixed temperature T using the Metropolis sampling technique [8]. The intermolecular potential is described by the sum of the Lennard-Jones and Coulomb potential

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

where Σ^a is the sum over the atoms of molecule a ; Σ^b is the sum over the atoms of molecule b ; r_{ij} is the distance between atoms i and j ; $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ and $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$; and σ_i , ϵ_i , and q_i are the potential parameters of atom i .

The average value of the energy is given as the average over a chain of size τ of instantaneous values:

$$\langle U \rangle = \frac{1}{\tau} \sum_i^\tau U_i, \quad (2)$$

and the error caused by the finite size of the simulation can be obtained as

$$\sigma(\langle U \rangle) = \sqrt{\frac{1}{\tau - 1} \langle \delta U^2 \rangle}, \quad (3)$$

where

$$\langle \delta U^2 \rangle = \frac{1}{\tau} \sum_i^\tau (U_i - \langle U \rangle)^2 = \langle U^2 \rangle - \langle U \rangle^2 \quad (4)$$

is the variance of U . Clearly, if τ approaches infinity, σ approaches zero.

The number τ of MC configurations necessary for the ensemble average can be reduced dramatically over the entire number of MC steps without modifying the calculated average energy. This is a consequence of the high correlation between successive MC configurations generated from simulations of dense systems such as liquids. The changes between successive configurations are small and cause high correlation between them. Using correlated configurations is, of course, a waste that does not contribute to the average. An analysis of the efficiency of the simulation shows how many MC steps are necessary to uncorrelate the configuration. This can be done by calculating the statistical inefficiency or the time (or step) correlation function [5, 7].

The statistical inefficiency adopted here is that defined by Friedberg and Cameron [9], and it is based on the mean-square deviations of averages taken over blocks of a chain of successive configurations. Dividing the chain of configuration in n_b blocks with t_b successive configurations, where $\tau = n_b t_b$, then the average taken over the b th block is

$$\langle U \rangle_b = \frac{1}{t_b} \sum_{i=b_{\text{start}}}^{b_{\text{end}}} U_i, \quad (5)$$

where $b_{\text{start}} = (b - 1)n_b + 1$, $b_{\text{end}} = bt_b$, and b can assume values from 1 to n_b . The mean-square deviation of averages taken over the blocks is

$$\langle \delta U^2 \rangle_b = \frac{1}{n_b} \sum_{b=1}^{n_b} (\langle U \rangle_b - \langle U \rangle)^2 \quad (6)$$

and the statistical inefficiency s is

$$s = \lim_{t_b \rightarrow \infty} \frac{t_b \langle \delta U^2 \rangle_b}{\langle \delta U^2 \rangle} = \lim_{t_b \rightarrow \infty} S(t_b), \quad (7)$$

where $\langle U \rangle$ is the average and $\langle \delta U^2 \rangle$ is the variance over the chain of the whole simulation as shown in Eqs. (2) and (4), while $\langle U \rangle_b$ is the average and $\langle \delta U^2 \rangle_b$ is the variance over the blocks. For chains of correlated configurations, $S(t_b)$ will increase with increasing t_b until it approaches a limiting value s . The limiting value of $S(t_b)$ signi-

fies that the block size t_b has become so large that there is no correlation from block to block. However, for finite chains ($\tau = n_b t_b$), as the block size t_b increases, the number of blocks n_b decreases and it generates an increased error in the calculation of $S(t_b)$ and, consequently, an imprecision in the estimate of the best value of s . Therefore, an analysis of s with respect to the size of the chain τ is very important.

For chains of uncorrelated configurations, Eq. (3) is substituted by

$$\sigma(\langle U \rangle) = \sqrt{\frac{1}{n-1} \langle \delta U^2 \rangle}, \quad (8)$$

where $n = \tau/s$. In that way, one can see that for a finite simulation the error of a property U is the same if one takes τ successive configurations or, equivalently, τ/s configurations separated by s MC steps. Thus, for a given simulation, obtaining ensemble averages over τ/s uncorrelated configurations is the same as over the entire number of configurations generated.

Results

Our working example to illustrate the favorable use of the statistical inefficiency is the solvatochromic shift of the first absorption $B_{2u}(\pi - \pi^*)$ band of benzene in water. We simulate [10] one benzene surrounded by 343 water molecules in the NVT ensemble at $T = 25^\circ$ and $\rho = 0.9966 \text{ g/cm}^3$. We use the potential parameters suggested by Jorgensen and coauthors: the SPC model for water [11] and the OPLS model for benzene [12] [$\epsilon_C = 0.110 \text{ kcal/mol}$, $\sigma_C = 3.750 \text{ \AA}$, $q_C = 0$, $\epsilon_H = 0$, $\sigma_H = 0$, $q_H = 0$, $r_{CC} = 1.400 \text{ \AA}$, $r_{CH} = 1.088 \text{ \AA}$, $\theta(\text{CCC}) = \theta(\text{CCH}) = 120^\circ$]. The cutoff radius is $r_c = 10.92 \text{ \AA}$ and we save 20,000 successive configurations.

To reduce the number of solvent molecules for the QM calculations, we use the RDF of the carbons of the benzene and oxygens of the water. The integration of the first peak gives the number of solvent molecules that compose the first solvation shell. In this case, the first shell has 18 water molecules.

All QM calculations reported next are made at the semi-empirical level using the ZINDO program [13]. The energy shift ΔE is obtained for each supermolecule using singly excited CI with

spectroscopic parametrization. In each configuration, only the benzene and the 18 water molecules that are the nearest neighbors will be considered as the supermolecule for our QM calculation.

In the statistical inefficiency analysis, we systematically use Eq. (7) to calculate the $S(t_b)$. In Figure 1, we show the behavior of $S(t_b)$ vs. t_b for chains of different sizes. We obtain the statistical inefficient, $s = 800 \pm 100$ MC steps. Now, based on this statistical inefficiency, we select from the 20,000 saved configurations a set of 25 configurations separated by 800 MC steps, e.g., a set of 25 uncorrelated configurations. To show that statistical information is not lost in this procedure, we calculate the ensemble average of the $B_{2u}(\pi - \pi^*)$ absorption transition of benzene in water for the entire simulation along with several sets of 25 uncorrelated configurations. For the entire simulation, we obtain a red shift compared to the gas phase of $\langle \Delta E \rangle = -122 \pm 15 \text{ cm}^{-1}$, which is in excellent agreement with the experimental shift of $\Delta E = -140 \text{ cm}^{-1}$ obtained by Bayliss and Hulme [14]. The calculated values of $\langle \Delta E \rangle$ for different sets of 25 uncorrelated configurations are shown in Table I. As we see, it not only produces the same results obtained above for the entire simulation but also it is irrelevant which particular set of 25 configurations is used. In Figure 2, we show the distribution of the ΔE_i obtained for each supermolecule used in the calculations of $\langle \Delta E \rangle$ presented in Table I. Finally, we used this approach successfully for the solvatochromism of benzene in several solvents and of acetone in water.

Conclusions

In this article, the statistical inefficiency of a Monte Carlo (MC) simulation is obtained. This can be used in the calculation of any thermodynamic property, giving a great reduction in the number of configurations necessary for the ensemble average. As the ensemble average is also necessary in the calculation of solvent effects, this reduction in the total number of expensive QM calculations can be made in a systematic way, without compromising the final average result.

ACKNOWLEDGMENTS

The authors thank Dr. Javier Satulovsky for helpful discussions. This work was supported by CNPq and FAPESP.

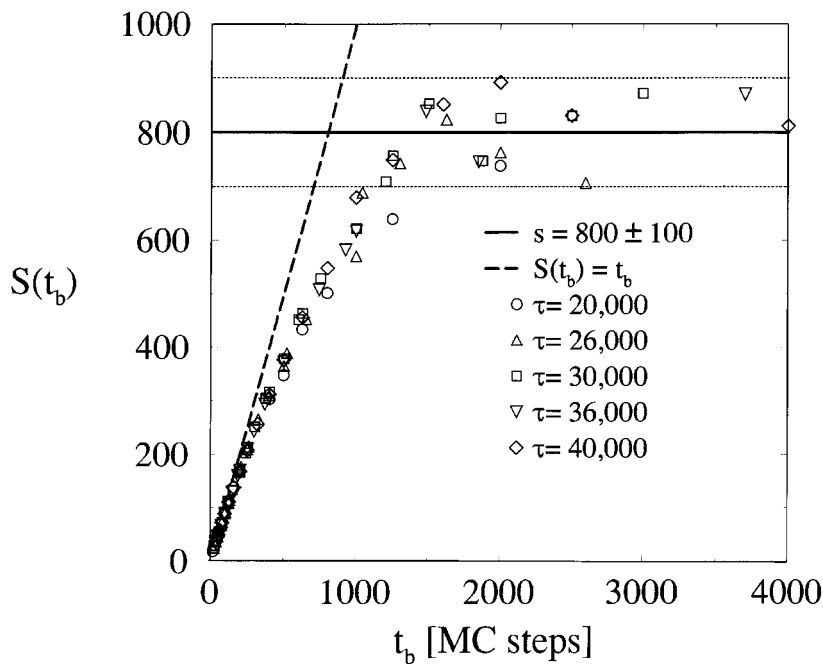


FIGURE 1. The behavior of $S(t_b)$ vs. t_b for chains of different sizes. The boldface lines represent the asymptotic behavior of correlated blocks (dashed) and uncorrelated blocks (solid). The estimated statistical inefficiency is $s = 800 \pm 100$ MC steps.

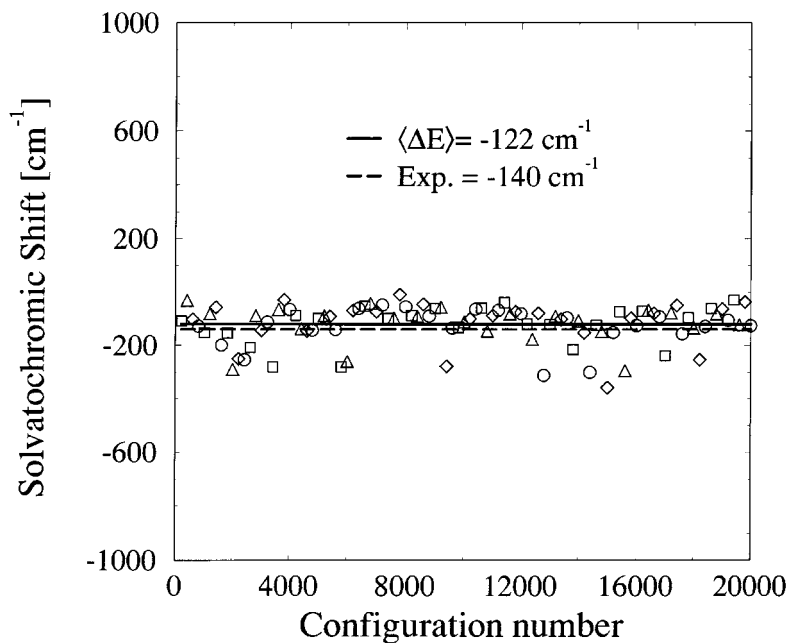


FIGURE 2. The distribution of spectral shift, ΔE_i , calculated with selected configurations. The solid line is the averaged spectral shift $\langle \Delta E \rangle = -122 \pm 15 \text{ cm}^{-1}$ and the dashed line is the experimental result [14], $\Delta E = -140 \text{ cm}^{-1}$. The different symbols represent sets 1–4 as shown in Table I.

TABLE I
Calculated shift of the $B_{2u}(\pi - \pi^*)$ transition of benzene in water.

Set	$\langle \Delta E \rangle$ (cm ⁻¹)
1	-124
2	-120
3	-114
4	-131

The displayed shifts are the average over 25 uncorrelated configurations (1 C₆H₆ + 18 H₂O). In all sets, the selected configurations are separated by 800 MC steps and the difference between them is the first selected configuration, e.g., set 1 is composed of the following configurations: 1, 801, 1601, 2401, ...; set 2: 201, 1001, 1801, ...; set 3: 401, 1201, 2001, ...; and set 4: 601, 1401, 2201,

References

1. C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry* (Verlag-Chemie, Weinheim, New York, 1979); E. S. Amis and J. F. Hinton, *Solvent Effects on Chemical Phenomena* (Academic Press, New York, 1973), Vol. 1; P. Beak, *Acc. Chem. Res.* **10**, 186 (1977); M. F. Nicol, *Appl. Spectrosc. Rev.* **8**, 183 (1974); N. S. Bayliss and E. G. McRae, *J. Phys. Chem.* **58**, 1002 (1954).
2. J. Tomasi and M. Persico, *Chem. Rev.* **94**, 2027 (1994); C. J. Cramer and D. G. Truhlar, in *Reviews in Computational Chemistry*, D. B. Boyd and K. B. Lipkowitz, Eds. (VCH, New York, 1995) Vol. 6; J. Gao, *Acc. Chem. Res.* **29**, 298 (1996); M. M. Karelson and M. C. Zerner, *J. Phys. Chem.* **96**, 6949 (1992).
3. J. T. Blair, K. Krogh-Jespersen, and R. M. Levy, *J. Am. Chem. Soc.* **111**, 6948 (1989); J. Gao, *J. Am. Chem. Soc.* **116**, 9324 (1994).
4. A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions* (Wiley, New York, 1991).
5. K. Coutinho and S. Canuto, *Adv. Quantum Chem.* **28**, 89 (1997).
6. F. J. Luque, J. M. Bofill, and M. Orozco, *J. Chem. Phys.* **103**, 10183 (1995); M. Orozco, F. J. Luque, D. Habibollahzadeh, and J. Gao, *J. Chem. Phys.* **102**, 6145 (1995).
7. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987).
8. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
9. R. Friedberg and J. E. Cameron, *J. Chem. Phys.* **52**, 6049 (1970).
10. K. Coutinho and S. Canuto, DICE: A Monte Carlo program for molecular liquid simulation (University of São Paulo, Brazil).
11. W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, *J. Chem. Phys.* **79**, 926 (1983).
12. W. L. Jorgensen, J. D. Madura, and C. J. Swenson, *J. Am. Chem. Soc.* **106**, 6638 (1984).
13. J. Ridley and M. C. Zerner, *Theor. Chim. Acta*, **32**, 111, (1973).
14. N. S. Bayliss and L. Hulme, *Aust. J. Chem.* **6**, 257 (1953).