Available online at www.sciencedirect.com
ScienceDirect
Chemical Physics Letters 437 (2007) 148–152
CHEMICAL PHYSICS LETTERS
www.elsevier.com/locate/cplett

An efficient statistically converged average configuration for solvent effects

K. Coutinho [a], H.C. Georg [a], T.L. Fonseca [a,b], V. Ludwig [a], S. Canuto [a,*]

[a] *Instituto de Física, Universidade de São Paulo, CP 66318, 05315-970, São Paulo, SP, Brazil*
[b] *Instituto de Física, Universidade Federal de Goiás, CP 131, 74001-970, Goiânia, GO, Brazil*

Received 17 November 2006; in final form 20 January 2007
Available online 9 February 2007

Abstract

Using statistically uncorrelated solute–solvent configurations generated by Monte Carlo simulation a simpler and efficient implementation of the averaged solvent electrostatic potential is made. An average configuration alone is used such that one single quantum mechanical calculation reproduces the converged statistical average obtained from the entire simulation. Applications are presented for solvent effects in a variety of properties of acetone and aminopurine in water. In all cases, excellent agreement is obtained using the average configuration and the average from the full statistical distribution.

1. Introduction

In recent years, we have seen considerable progress in the study of liquid systems and solvent effects [1–5]. This is motivated by the great importance of solvents in several different aspects of molecular structure, spectra and reactivity. The idea of combining molecular mechanics (MM) and quantum mechanics (QM) originated the so-called QM/MM methods [3–5]. One variant is the use of computer simulation to generate the structure of the liquid for subsequent QM calculations [6,7]. The advantage of performing the calculations sequentially is that after the simulation statistical information permits an efficient protocol for the QM calculations. In this way, statistically converged results can be obtained with a relatively small number of the usually expensive QM calculations. In general, we have been able to obtain statistically converged average results with less than 100 QM calculations. If, for one hand this is a relatively small number, compared to conventional on-the-fly QM/MM methodologies, that may require thousands of QM calculations, it is still large enough to preclude the incursion into large molecules. For molecules of biological interest, or for high-level computational-demanding calculations, for instance, this could still be a severe limitation. Therefore, it would be very desirable to overcome this bottleneck reducing further the necessary number of QM calculations to obtain any average property. In more dramatic terms it would be desirable to perform *just one* QM calculation. One possibility would be to have an average solute–solvent potential that could reproduce the collection of available configurations, even if unphysical, but reproducing the average of any property involved. In fact, this possibility has been developed by Aguilar and co-workers in the form of an averaged solvent electrostatic potential (ASEP) [8–10]. In this work, we explore this making a simpler and efficient implementation using statistically uncorrelated solute–solvent configurations generated by Monte Carlo (or molecular dynamics) simulation. If the liquid solvent around the solute can be represented by classical point charges the applications below will show that the use of an average configuration alone is sufficient to obtain the average value, though at the expenses of the statistical distribution. Applications will be made for acetone and aminopurine in water, using var-

* Corresponding author. Fax: +55 11 3091 6831.
E-mail address: canuto@if.usp.br (S. Canuto).

0009-2614/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.cplett.2007.02.012

ious theoretical models and properties, including the NMR shieldings in the ^{17}O and ^{13}C atoms of acetone that have attracted recent interest [11,12]. To obtain these properties, we first use the usual average procedure using statistically uncorrelated configurations [6,7]. Next, we use these configurations to generate the average configuration and obtain the same properties with just one QM calculation. As we will see they give the same numerical values, thus opening the possibility to tackle real large systems without compromising the statistical average.

2. Calculation details

2.1. Monte Carlo and quantum mechanical calculations

The MC simulations were performed using the Metropolis sampling technique at normal temperature [13]. The simulation of acetone in water consisted of one polarized acetone molecule embedded in $N = 450$ molecules of water as described before [14]. After thermalization, 6.57×10^7 MC steps are performed. Calculating the statistical interval obtained from the auto-correlation function of the energy [7], we have sampled $M = 100$ configurations for the QM calculations with less than 10% of statistical correlation. These were composed of the central acetone plus $L = 200$ water molecules (corresponding to the cut-off radius of 11.5 Å). For aminopurine in water the simulation has also been described before [15]. In this case, we separated $M = 45$ configurations composed of the central aminopurine plus $L = 456$ water molecules. In all cases, water is represented by simple-point charges (SPC model) in the MC simulation and by atomic charges in the QM calculations. A variety of QM models are used for different properties. For acetone in water, the dipole moments are calculated using MP2/aug-cc-pVDZ, the excitation energies are calculated using time-dependent DFT in the hybrid B3LYP/6-311++G(d,p) model and the NMR shielding parameters $\sigma^{iso}(^{13}C)$ and $\sigma^{iso}(^{17}O)$ are obtained also with DFT but using GIAO and an improved basis set, B3LYP/6-311++G(2d,2p). For aminopurine, we calculate dipole moments using CASSCF(12,10). For the excitation energies they were supplemented with second-order perturbation theory CASPT2. Basis set was the atomic natural orbital of triple-zeta plus polarization supplemented by 1s1p1d Rydberg functions [15]. QM calculations used the MOLCAS [16] for CASSCF and CASPT2 and GAUSSIAN 98 [17] for MP2 and B3LYP. The MC simulations used the DICE program [18].

2.2. The statistically converged average configuration

We use the idea of the ASEP developed by Aguilar and co-workers [8–10] but we make a simpler implementation. Because the configurations separated here from the simulations are statistically uncorrelated, a relatively small number of them should give convergent results. Thus we simply superpose the coordinates of all L solvent molecules (with scaled charges), within a given cut-off radius, in the M configurations, with the solute fixed. This is the ASEP configuration [8] but we use it here for all solvent molecules (L) within a large cut-off radius. For large enough cut-off the long-range contribution is included. Hence, we avoid generating a fitting of an average potential [8–10]. In this very simple implementation only the averaged configuration is used, with no additional procedures, making the applications straightforward.

The statistical interval obtained from the auto-correlation function of the energy, $C(i)$, is very important. For a markovian process, $C(i)$ follows an exponential decay [7], as seen in Fig. 1 for the case of acetone in water. From that we can select $M = 100$ configurations that have less than 10% of statistical correlation. These 100 configurations are used in the QM calculations to obtain the averages and the statistical distribution. For the average configuration we simply superpose all these M configurations of solvent atomic charges, and scale it by $1/M$. We thus have *one configuration* composed by the solute surrounded by $M \times L$ solvent molecules represented by atomic charges of values scaled by $1/M$. The scaling is important for giving the proper normalized average configuration. For example, in the case of acetone instead of performing 100 QM calculations, each one composed by 1 acetone surrounded by 600 point charges (i.e. 200 water molecules with charge q_O in the oxygens and q_H in the hydrogens), we perform only one QM calculation using the configuration composed by 1 acetone surrounded by 60 000 point charges (i.e. 200 water molecules on 100 superposed configurations with charges $q_O/100$ and $q_H/100$). The same procedure is used for aminopurine with $M = 45$ and $L = 456$. This is an average single configuration extending to the outer layers of solvent and ensuring statistically convergent results, avoiding separate consideration of the long range potential. Only the average configuration is used and hence in the following we shall term this as ASEC.

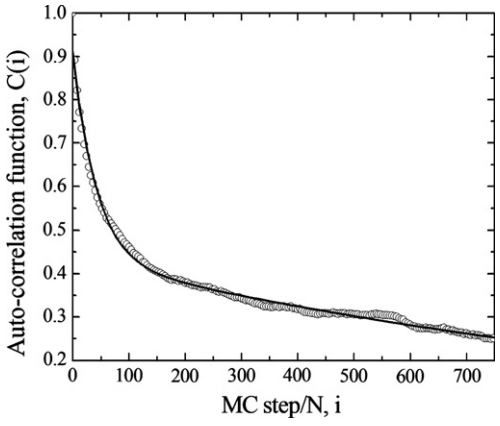

Fig. 1. Auto-correlation function of the energy for acetone in water and best fit of $C(i) = 0.47e^{-(i/30)} + 0.48e^{-(i/950)}$. For an interval of 6.75×10^5 MC steps, $i = 6.75 \times 10^5/450 = 1500$ and $C(1500) = 0.099$, giving a statistical correlation of 9.9%.

Table 1
Average values obtained with the average configuration (ASEC) and with 100 uncorrelated configurations (100 QM) for acetone in water

	μ	n–π^*shift	$\Delta\sigma^{iso}(^{13}C)$	$\Delta\sigma^{iso}(^{17}O)$
	MP2/aug-cc-pVDZ	TD-B3LYP/6-311++G(d,p)	B3LYP/6-311++G(2d,2p)	
ASEC	4.79	0.31	−20.5	120.5
100 QM	4.80 ± 0.03	0.31 ± 0.01	−20.5 ± 0.4	120.3 ± 2.1

Dipole in Debye, energy in eV and NMR shieldings in ppm.

Table 2
Average values obtained using the average configuration (ASEC) and 45 uncorrelated configurations (45 QM) for the N9H tautomer of aminopurine in water

	μ (D) CASSCF		ΔE (eV) CASPT2
	Ground state	Excited state π–π^*	π–π^* transition
ASEC	5.57	6.94	4.02
45 QM	5.60 ± 0.12	6.92 ± 0.21	4.02 ± 0.02

3. Results and discussion

The numerical results are shown in Tables 1 and 2 and will be discussed individually in the following sub-sections.

3.1. Acetone in water

3.1.1. Dipole moment and solvatochromic shift

The average in-solution dipole moment of acetone in water is 4.80 ± 0.03 D, obtained from 100 QM calculations using MP2/aug-cc-pVDZ. This value is in very good agreement with the Car–Parrinello MD result of 4.90 D [19]. Table 1 shows the result of 4.79 D obtained using only ASEC. The agreement for the average values obtained by the two procedures is excellent. Good agreement for dipole moments has also been obtained using ASEP [20].

Table 1 shows the calculated n–π^* transition shift using the average of 100 uncorrelated configurations as 0.31 ± 0.01 eV (~2500 cm^{-1}). This value is slightly larger than the experimental result (see Refs. [14,21]), between 1500 and 1700 cm^{-1} but is in agreement with several previous results obtained by representing the solvent as point charges (see Ref. [21]). We attribute this to the absence of dispersion interaction that is known to decrease (contribute to a red shift) the solvatochromic shift [22]. In any case, again, the result for the shift is in excellent accord with the single calculation using ASEC (Table 1).

3.1.2. NMR shielding parameters $\sigma^{iso}(^{13}C)$ and $\sigma^{iso}(^{17}O)$

The solvent contribution to nuclear shielding constants is $\Delta\sigma = \sigma^{solution} - \sigma^{vacuo}$. We then calculated the isotropic constants $\sigma^{iso}(^{13}C)$ and $\sigma^{iso}(^{17}O)$ of acetone in vacuo and in water. In both cases, the two constants are negative (deshield) but the aqueous shifts are in opposite directions. $\Delta\sigma^{iso}(^{17}O)$ is positive whereas $\Delta\sigma^{iso}(^{13}C)$ is negative. As

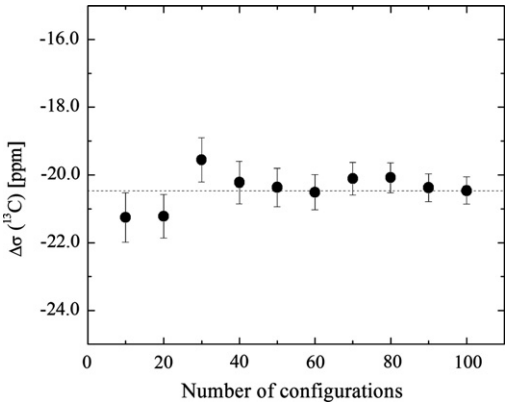

Fig. 2. Statistical convergence of the average solvent shift of $\sigma^{iso}(^{13}C)$ for acetone in water.

Table 1 shows the solvent effects on these shieldings are $\Delta\sigma^{iso}(^{13}C) = -20.5 \pm 0.4$ ppm and $\Delta\sigma^{iso}(^{17}O) = 120.3 \pm 2.1$ ppm. Experimental result [23] for the shift in ^{13}C is −18.5 ppm, in very good agreement with our calculated result. For ^{17}O our result is larger than the available experimental result [11] of 75.5 ppm. Two aspects should be considered. One is the effect of nuclear vibrations. Mennucci et al. [24] conclude that vibrations do not affect much the average value of NMR shieldings. The other is the so-called indirect effects that decrease the direct estimate of the shieldings [12]. Compatibly, our result is in very good agreement with the recent direct result of 121.8 ppm obtained using Car–Parrinello MD [12]. But the main objective here is the performance of the ASEC. It is again seen that *one QM calculation* using ASEC reproduces precisely the statistical average using all configurations selected. Fig. 2 shows the statistical convergence of $\Delta\sigma^{iso}(^{13}C)$ using all 100 uncorrelated structures, and clearly shows that the same value obtained with ASEC thus represent the converged value. The use of ASEC produces the same average values but suppresses the statistical distribution. This is illustrated in Fig. 3, that shows the histogram

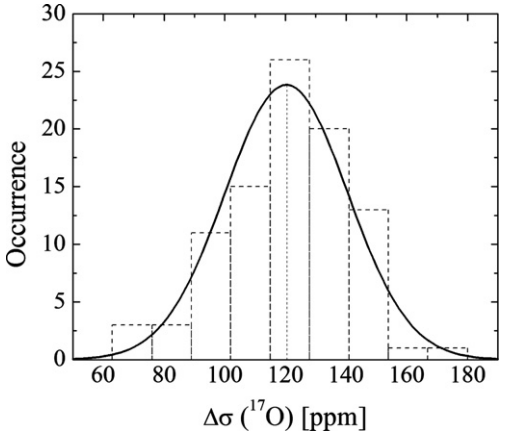

Fig. 3. Statistical distribution of the calculated $\Delta\sigma^{iso}(^{17}O)$ and the average value obtained using ASEC (central dotted line) for acetone in water.

of the statistical distribution of calculated values for $\Delta\sigma^{iso}(^{17}O)$ compared to the single average value obtained with ASEC.

3.2. Aminopurine in water

As an example of a computational-demanding problem we now consider aminopurine in water using the high-level CASSCF and CASPT2 theoretical models. This is the situation where the use of a single structure, obtaining the statistical average from just one QM calculation, opens wider possibilities. We will consider the calculations of the in-solution dipole moment and the solvent effect on the lowest π–π^* transition. Aminopurine consists of six- and five-member rings, thus modeling several molecules of great interest such as guanine, adenine, indole, etc. We consider the N9H tautomer that is the more stable species in water [25].

3.2.1. Dipole moment and solvatochromic shift

Table 2 shows the calculated dipole moment of aminopurine in water both for the ground and the first excited π–π^* state, obtained using state-average CASSCF. The gas phase ground state dipole moment is calculated as 3.02 D increasing to 5.60 ± 0.12 D in the water environment. The excited state gas phase is calculated as 3.51 D increasing to the value of 6.92 ± 0.21 D in water. As the dipole moment increase upon excitation we expect the corresponding excitation to decrease in water. In fact, the gas phase excitation is calculated as 4.07 eV decreasing to 4.02 eV in water, giving a red shift of 0.05 eV. The calculated excitation energy of 4.02 eV is in very good agreement with the experimental absorption measured in aqueous solution between 4.05 and 4.08 eV [26,27]. Again, Table 2 shows that the single QM calculation using ASEC gives the same averages as that obtained using all 45 configurations. Fig. 4 confirms that these correspond to statistically converged results.

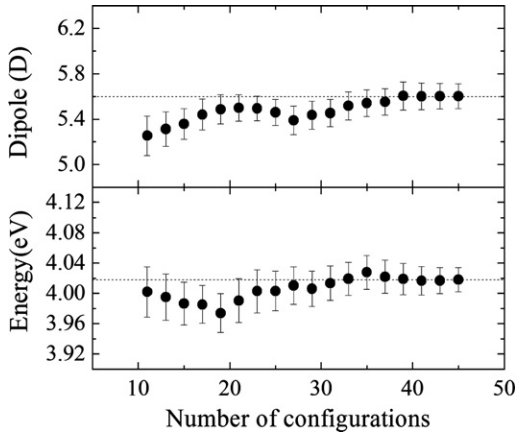

Fig. 4. Statistical convergence of the ground state dipole moment (top) and the excitation energy (bottom) to the low-lying π–π^* of aminopurine in water.

4. Summary and conclusions

Using statistically uncorrelated solute–solvent configurations generated by Monte Carlo (or molecular dynamics) simulation a simpler and efficient implementation of the averaged solvent electrostatic potential (ASEP) is made. An average configuration alone (ASEC) is used such that one single quantum mechanical calculation reproduces the converged statistical average obtained from the entire simulation. This is a straightforward implementation that excludes fitting an average potential, and all its subsequent uses.

Test cases are presented for in-solution dipole moments and shifts of n–π^* and π–π^* transitions. Additional application is made to obtain the NMR shielding parameters $\sigma^{iso}(^{13}C)$ and $\sigma^{iso}(^{17}O)$ of acetone in water. These exemplify molecular properties commonly of interest in the study of solvent effects. In all cases, excellent agreement is obtained for the averages using the average configuration and the full statistical distribution. Widely used theoretical models, such as MP2, B3LYP, CASSCF and CASPT2, have shown the validity and the possibility of vast applications. This corroborates the importance of the ASEP methodology [8–10]. It is not expected to be used, in its present form, in situations where the electronic structure of the solvent molecules are necessary but it is very effective when the solute molecule is placed in the point charges of the solvent molecules. As it replaces the statistical distribution for just one average configuration it suppresses the statistical distribution. Hence one may not apply this to study inhomogeneous broadening and line width of the absorption spectra of liquid systems, for instance. But it is valid for efficiently obtaining average values in electrostatic fields, a situation that is most common in theoretical studies of solvent effects.

Acknowledgements

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

References

[1] C.J. Cramer, Essentials of Computational Chemistry, in: Theories and Models, Wiley, New York, 2002.
[2] J. Tomasi, B. Mennucci, R. Cammi, Chem. Rev. 105 (2005) 2999.
[3] M. Orozco, F.J. Luque, Chem. Rev. 100 (2000) 4187.
[4] M.F. Ruiz-López, editor, J. Mol. Struct. (Theochem) 632 (2003), special issue on 'Combined QM/MM calculations in chemistry and biochemistry', p. 1.
[5] J. Gao, in: K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 7, VCH, New York, 1996, p. 119.
[6] K. Coutinho, S. Canuto, Adv. Quantum Chem. 28 (1997) 89.
[7] K. Coutinho, S. Canuto, J. Chem. Phys. 113 (2000) 9132.
[8] M.L. Sánchez, M.A. Aguilar, F.J. Olivares Del Valle, J. Comput. Chem. 18 (1997) 313.
[9] M.L. Sánchez Mendoza, M.A. Aguilar, F.J. Olivares Del Valle, J. Mol. Struct. (Theochem) 426 (1998) 181.
[10] M.E. Martin, M.L. Sánchez, F.J. Olivares Del Valle, M.A. Aguilar, J. Chem. Phys. 113 (2000) 6308.
[11] M. Cossi, O. Crescenzi, J. Chem. Phys 118 (2003) 8863.

[12] M. Pavone, O. Crescenzi, G. Morelli, N. Rega, V. Barone, Theor. Chem Acc. 116 (2006) 456.
[13] M.P. Allen, D.J. Tildesley, Computer Simulation of Liquids, Clarendon Press, Oxford, 1987.
[14] H.C. Georg, K. Coutinho, S. Canuto, Chem. Phys. Lett. 429 (2006) 119.
[15] A.C. Borin, L. Serrano-Andrés, V. Ludwig, K. Coutinho, S. Canuto, Int. J. Quantum Chem. 106 (2006) 2564.
[16] K. Andersson et al., MOLCAS version 5, Lund University, Lund, Sweden, 2000.
[17] M.J. Frisch et al., GAUSSIAN 98 (Revision A.7), Gaussian, Inc., Pittsburgh, PA, 1998.
[18] K. Coutinho, S. Canuto, DICE: A Monte Carlo program for molecular liquid simulation, University of São Paulo, São Paulo, 2003.
[19] U.F. Röhrig, I. Frank, J. Hutter, A. Laio, J. Vandevondele, U. Rothlisberger, ChemPhysChem 4 (2003) 1177.
[20] M.L. Sánchez, M.E. Martin, I. Fdez. Galván, F.J. Olivares Del Valle, M.A. Aguilar, J. Phys. Chem. 106 (2002) 4813.
[21] K. Aidas, J. Kongsted, A. Osted, K.V. Mikkelsen, O. Christiansen, J. Phys. Chem. A 109 (2005) 8001.
[22] S. Canuto, K. Coutinho, M.C. Zerner, J. Chem. Phys. 112 (2000) 7293.
[23] B. Tiffon, J.E. Dubois, Org. Magn. Reson. 11 (1978) 295.
[24] B. Mennucci, J.M. Martinez, J. Tomasi, J. Phys. Chem. A 105 (2001) 7287.
[25] C.E. Crespo–Hernández, B. Cohen, P.M. Hare, B. Kohler, Chem. Rev. 104 (2004) 1977.
[26] A. Holmén, B. Nordén, B. Albinsson, J. Am. Chem. Soc. 119 (1997) 3114.
[27] S.K. Pal, J. Peon, A.H. Zewail, Chem. Phys. Lett. 363 (2002) 57.