Canuto, Coutinho and Trzesniak, Adv. Quantum Chem. 41, 161 (2002).

New Developments in Monte Carlo/Quantum Mechanics Methodology. The Solvatochromism of β-Carotene in Different Solvents.*

by Sylvio Canuto[1], Kaline Coutinho[2] and Daniel Trzesniak[1]

[1]Instituto de Física, Universidade de São Paulo, CP 66318
05315-970 São Paulo, SP, Brazil.

[2]Universidade de Mogi das Cruzes/CCET, CP 411
08701-970, Mogi das Cruzes, SP, Brazil.

March 28, 2003

Abstract

The solvatochromic shifts of the $\pi - \pi^*$ transition of all-*trans*-β-carotene in isopentane, acetone, methanol and acetonitrile are studied using a sequential Monte Carlo/quantum mechanics (S-MC/QM) methodology. These different solvents are examples of systems of varied nature, differing in dielectric constants and covering a wide range of polarities, and including also polar and non-polar solvents. In S-MC/QM we first generate the structure of the liquid using Metropolis MC simulation and then perform the QM calculations in statistically uncorrelated configurations. It is shown that, in these cases, including only 40 QM calculations gives statistically converged results. To deal with elongated solutes the box of the MC simulation has been extended to a large rectangular shape. Then, a nearest-neighbor distribution function has been developed and generalizes the concept of solvation shells for a solute of any arbitrary shape. The calculated results are converged with respect to the number of solvent molecules that are included according to the nearest-neighbor distribution function. The results are found to be in very good quantitative and qualitative agreement with experiment. The dipole moments of the ground and excited $\pi - \pi^*$ states of β-carotene

*Published in *Adv. Quantum Chem.* **41**, 161 (2002).

are both zero and the transition shifts are thus dominated by the dispersive interaction. The inclusion of dispersion interaction in energy differences is then discussed.

Contents

1 Introduction

The study of molecular systems in the liquid phase is important for understanding a great number of chemical, physical and biological processes[1]. The solvent interaction leads to changes in the molecular solute affecting its spectroscopic, structural and reactive properties. For this reason, the study of solvent effects has been a topic of increased interest[2, 3, 4]. In the theoretical front the basic ideas developed by Onsager[5] and Kirkwood[6] have led to sophisticated cavity theories, where the solute is enclosed in a cavity and the solvent is treated by a continuum polarizable dielectric medium. Tapia and Goscinski[7] have developed one of the first successful self-consistent reaction field (SCRF) theories that has been extended further by many others[8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. Present continuum models include sophisticated procedures, where the solute is treated with electron correlation effects[15, 16] leading to more accurate reaction fields, and variants such as the COSMO[12] methodology. Warshel and Levitt[18] have suggested a hybrid quantum mechanical-molecular mechanics (QM/MM) methodology, where the most important part of the system is treated by quantum mechanics and the rest by classical mechanics. Thus in solute-solvent interaction the chromophore, and perhaps a few other molecules, are treated by QM and the solvent is considered by classical point charges[19, 20, 21, 22, 23, 24, 25]. This idea was further developed by Blair and co-workers[25], by Gao[20] and by Zeng and co-workers[26] that

considered that a liquid has not one but many structures at a certain temperature. A liquid is, indeed, statistical by nature and the liquid properties are, in fact, statistical averages. Thus they performed Molecular Dynamics and Monte Carlo simulations to generate the structure of the liquid. Gao has further developed this idea generating a successful Monte Carlo QM/MM method[20, 21, 22].

In Monte Carlo simulation of liquids the configurational space necessary for configurational averages is generated by Metropolis sampling technique and includes temperature effects. Although this is a more realistic representation of the liquid nature of the solvent, it has the concomitant disadvantage that several quantum mechanical calculations are necessary to obtain the proper statistical average. For instance, in studying solvatochromic shifts the transition energy has to be calculated several times for structures generated by the simulation, in order to obtain the average value that corresponds to the solvation shift. In many cases millions of calculations have been performed on these supermolecular systems composed of the solute treated by QM and the solvent as classical point charges. Furthermore, if the solvent is not explicitly treated by QM it is difficult to include dispersion interaction that, in fact arises from the reciprocal polarizations of the solute by the solvent, and the solvent by the solute.

We have extended this idea to a sequential Monte Carlo/Quantum mechanics (S-MC/QM)[27]. In this procedure we first generate structures of the liquid and only subsequently perform the QM calculations in those structures. The basic advantage is that opposite to conventional QM/MM, in the S-MC/QM, the solute and *all solvent molecules*, up to a certain solvation shell, are treated by quantum mechanics[28, 29]. The number of necessary solvation shells to be included can be systematically analyzed and converged results obtained[28]. As an important development, we have also shown that the drawback of having to perform a large number of quantum mechanical calculations to obtain the average of the property of interest may be strongly alleviated considering the statistical correlation between successive configurations[27, 29, 30]. As MC generates structures that belong to a markovian chain, the auto-correlation function of the energy gives important information on the relative statistical importance of the successive structures generated by the simulation. Of course, highly correlated structures will give very little new statistical information[30]. In other words, performing calculations on every structure generated is an enormous waste that gives no new statistical information. Reducing the number of quantum mechanical calculations is a great saving in computational resources that can be used to explicitly include the solvent molecules and thus leading to a more realistic treatment of the intermolecular solute-solvent interactions. As an example, the solvatochromic shift of pyrimidine in water[31] was suc-

cessfully treated including all water molecules up to the third solvation shell. This required supermolecular quantum-mechanical calculations of the pyrimidine and 213 water molecules, an explicit 1734 all-valence electrons, properly anti-symmetrized. With this procedure specific interactions such as charge transfer and hydrogen bonds are naturally treated. Detailed analysis of the convergence of the average value with the number of configurations are made elsewhere[30, 32], and shows that the average value is indeed converged.

Inclusion of dispersive interaction in solvent effects[33] has been a real challenge for present theoretical methodologies[2, 3]. If the solvent molecules are not explicitly included the polarization of the solute onto the solvent is not considered and dispersion is omitted. As dispersion is a double excitation, derived from single excitation in the solute and single excitation in the solvent, one possibility is to have previously calculated and separated the spectrum of the solvent molecule and try including this information in the calculation of the solvatochromic shift. This was attempted by Rösch and Zerner[34]. Another possibility is, of course, to explicitly include the solvent molecules in the supermolecular QM calculations. Our S-MC/QM procedure allows this to be done in a natural way[27, 28, 29]. In this connection it is very important to note that it is possible to include dispersive interaction in transition energy, using a singly-excited configuration interaction (CIS). It has been demonstrated[35] before that a configuration interaction electronic structure calculation on a supermolecule that contains only single excitations includes dispersion interactions between the two subsystems when energy differences are taken between the Hartree Fock (SCF) ground state and low energy excited states in which single excitations dominate. This theorem is proven up to second order in perturbation theory[35]. This has been used to calculate the solvatochromic shifts of benzene in different solvents (polar and non-polar)[29] with very good agreement with the experimental results. As the dipole moment of benzene is zero in the ground state and in the low-lying $\pi - \pi^*$ excited states the solvatochromic shifts in different solvents is basically given by dispersion (quadrupolar interaction is very small) and leads to a red shift, as described earlier by Liptay[36].

In short, our S-MC/QM methodology uses structures generated by MC simulation to perform QM supermolecular calculations of the solute and all the solvent molecules up to a certain solvation shell. As the wave-function is properly anti-symmetrized over the entire system, CIS calculations include the dispersive interaction[35]. The solvation shells are obtained from the MC simulation using the radial distribution function. This has been used to treat solvatochromic shifts of several systems, such as benzene in CCl_4, cyclohexane, water and liquid benzene[29, 37]; formaldehyde in water[28, 38]; pyrimidine in water and in CCl_4[31]; acetone in water[39]; methyl-acetamide in water[40] etc.

In this paper we address to the solvation of all-*trans*-β-carotene in different solvents. The solvent effects on the visible spectrum of β-carotene is a real challenge for theoretical methodologies for at least two aspects. First, the visible spectrum is characterized by a strong $\pi - \pi^*$ absorption transition in the region of 450 nm that suffers only slight shifts in different solvents[41, 42]. As the dipole moment is zero both in the ground and excited state, the shift is dominated by dispersion interaction. This interaction is, in addition, small for different solvents. For instance, the shift of the $\pi - \pi^*$ absorption transition from isopentane (non-polar, non-protic, has small polarity and small dielectric constant) to methanol (polar, protic, has large polarity and large dielectric constant) is only 120 cm^{-1}[42]. Because of the low volatility of β-carotene the gas phase value of the absorption transition is not known experimentally and correlation of solvatochromic shifts in different solvents is very important. Second, the elongated shape of the molecule (see below) suggests the use of a non-spherical distribution of solvent molecules around the solute. This is parallel to the cavity-shape problem in SCRF methods[43]. For our purposes, the use of spherically defined solvation shells is not recommended. This is a delicate point as the concept of solvation shells is in essence, but not compulsory, related to a spherical distribution[44]. We will see that this is not only inconvenient but, to some extent, incorrect. Thus, we develop a nearest-neighbor distribution that follows the molecular shape and can be used for any molecule, no matter how elongated or distorted. The visible spectrum of β-carotene has been analyzed before by Applequist[45] using a cavity model where the chromophore was treated as classical point dipole oscillator. Myers and Birge[41] studied the change in oscillator strength of the absorption of β-carotene in different solvents and found that the results depend on the prolate cavity geometry. Zerner made an estimate of the shift in cyclohexane[46] using SCRF. Abe and co-workers[42] analyzed solvent effects in 51 different solvents and made an empirical analysis in terms of reaction field models. Here, we use our S-MC/QM methodology in a nearest-neighbor solvation shell to calculate the solvatochromic shift of β-carotene in four different solvents; namely, isopentane, acetonitrile, acetone and methanol. These four solvents are selected on the basis of their nature, exemplifying polar, non-polar, protic, non-protic, low polarity and large polarity solvents.

2 Monte Carlo Simulation

2.1 Rectangular Box and Computational Details

Monte Carlo (MC) simulations were carried out for all-*trans*-β-carotene in four solvents: acetone ($(CH_3)_2CO$), acetonitrile (CH_3CN), isopentane

($(CH_3)_2CHCH_2CH_3$) and methanol (CH_3OH). Standard procedures[47] were used, including the Metropolis sampling technique[48] in the canonical (NVT) ensemble and periodic boundary conditions using the image method. Because of the prolate shape of β-carotene we used a rectangular box instead of the more conventional cubic box. Therefore, a cuttoff radius was not used, but each molecule was restricted to interact either with a molecule or its respective image, not simultaneously with both. Therefore, this system (1 solute + N solvent molecules) corresponds to an infinitely dilute solution. In figure 1, the solute (all-*trans*-β-carotene) in the smallest rectangular box used in our sim-

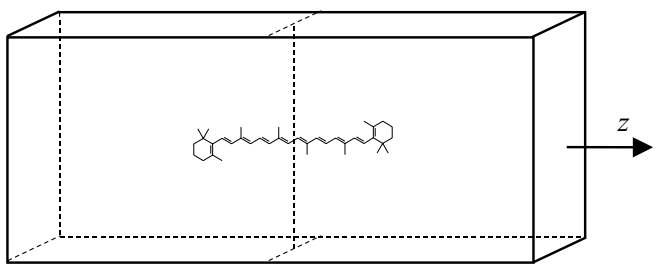

Figure 1: Illustration, in scale, of the all-*trans*-β-carotene in the smallest rectangular box (30x30x70.6)$\mathring{A}^3$ used in our simulations.

ulations is illustrated. Note that the β-carotene is almost a planar molecule with approximately 29 $\mathring{A}$ in the long axis and 6 $\mathring{A}$ in the small axis. Thus, even in the smallest box there was sufficient space to wrap the β-carotene in a bulk environment.

Table 1
Information of the simulated systems: the density, the box size, the dielectric constant and the polarity of the solvent.

Solvent	Density g/cm^3	Box Size (x, y, z) in $\mathring{A}$	Dielectric Constant (ϵ)	Normalized Polarity (E_T^N)
Isopentane	0.6001	$(45.5, 45.5, 87.5)$	1.828	0.006
Acetone	0.7682	$(38.5, 38.5, 77.0)$	21.36	0.355
Methanol	0.7676	$(30.0, 30.0, 70.6)$	32.66	0.762
Acetonitrile	0.7649	$(33.5, 33.5, 72.5)$	35.94	0.460

The four systems were simulated at $T = 298$K and were consisted of one β-carotene molecule and 900 solvent molecules in a rectangular box with linear dimensions, which correspond to the solvent densities[49]. The solvent density, the box size, the dielectric constant and the normalized E_T^N Reichardt polarity[1] are shown in table 1. Note the variation of the dielectric constants

of the selected solvents. These solvents also exhibit large variations of the normalized solvent polarity, changing from 0.006 (isopentane) to 0.762 (methanol) with the intermediate values of 0.355 (acetone) and 0.460 (acetonitrile)[42].

The intermolecular interactions were described by the Lennard-Jones plus Coulomb potential,

$$U_{ab} = \sum_{i}^{\text{on}\,a} \sum_{j}^{\text{on}\,b} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{e^2}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \tag{1}$$

where $\sum^a$ is the sum over the sites of molecule a, $\sum^b$ is the sum over the sites of molecule b, $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$, $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$, $e^2/(4\pi\epsilon_o)$= 331.9684 $\AA$ kcal/mol and σ_i, ε_i and q_i are the parameters of the interacting sites. The potential parameters of the sites used in the our simulations were obtained in the OPLS force field[50] and are shown in table 2. The geometry of the β-carotene, shown in figure 2, was obtained by gradient optimization, starting from the crystallographic experimental data[51], with the Becke three-parameter-functional[52] and the Lee-Yang-Parr correlation[53], B3LYP/6-31G level of calculation. The

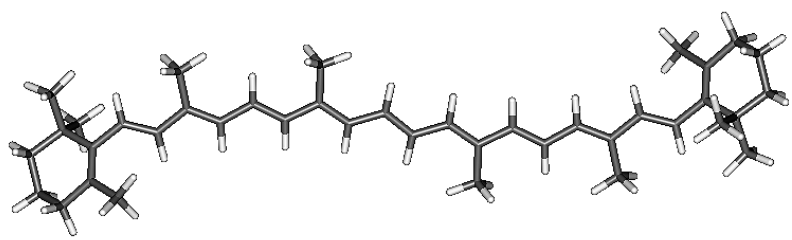

Figure 2: Geometry of the β-carotene obtained with B3LYP/6-31G optimization.

geometries of the solvents were obtained in the OPLS force field (acetone[54], acetonitrile[55], isopentane[56] and methanol[57]). All molecules were kept in the equilibrium geometry during the simulation. The initial configurations were generated randomly, considering the position and orientation of each molecule. A new MC step was generated by randomly selecting a solvent molecule, translating it randomly in all Cartesian directions and rotating it randomly about a randomly chosen axis. A new configuration was generated after 900 MC steps, i. e., after closing a loop over the solvent molecules. In this way, the number of configurations l generated here is equivalent to the number of configurations generated in a Molecular Dynamics simulation with an integration over l time steps. The acceptance of each random move was governed by the Metropolis sampling technique. The maximum displacement of the molecules was self adjusted after 50 MC steps to give an acceptance rate around 50%. The full simulation consisted of a thermalization stage of 4.5x10^6

Table 2
Potential parameters used in the Monte Carlo simulations (q_i in elementary charge unit, ε_i in kcal/mol and σ_i in $\mathring{A}$).

Site	q_i	ε_i	σ_i
Isopentane			
CH_3	0.000	0.160	3.910
CH_2	0.000	0.118	3.905
CH	0.000	0.080	3.850
Acetone			
O	−0.424	0.210	2.960
C	0.300	0.105	3.750
CH_3	0.062	0.160	3.910
Methanol			
H	0.435	0.000	0.000
O	−0.700	0.170	3.070
CH_3	0.265	0.207	3.775
Acetonitrile			
N	−0.430	0.170	3.200
C	0.280	0.150	3.650
CH_3	0.150	0.207	3.775
β-Caroteno			
$C(sp^2)$	0.000	0.105	3.750
$C(sp^3)$	0.000	0.050	3.800
CH_3	0.000	0.175	3.905
CH_2	0.000	0.118	3.905
CH	0.000	0.115	3.800

MC steps, which is not used in the statistics, followed by an averaging stage of 36x10^6 MC steps. This is a long simulation by all present standards. Thus, the total number of configurations generated in each MC simulation was $l = 40000$. Instead of performing a quantum mechanical calculation on every configuration generated by the Monte Carlo simulation, we use the auto-correlation or statistical efficiency, to select the statistically relevant structures[27, 29, 30, 32]. In doing so, the subsequent quantum mechanical calculations are performed only on some uncorrelated structures. As in previous works[28, 29, 31, 38, 40] we fit the auto-correlation function of the energy to an exponentially decaying function and obtain the correlation step. This assures that the structures used in the quantum mechanical calculations are statistically (nearly) uncorrelated.

As the total number of MC configurations generated in the simulation was 40000, the averages are then taken over only 40 configurations, separated by 1000 successive configurations. The convergence of the calculated values using this reduced number of uncorrelated configurations is discussed later in this paper. All simulation were performed with the DICE[58] Monte Carlo statistical mechanics program.

To obtain the relative solvatochromic shifts, the excitation energies were calculated using the ZINDO program[59], within the INDO/CIS[60] approach. The quantum mechanical calculations were performed for the supermolecular clusters, generated by the MC simulations, composed of one β-carotene and all solvent molecules within a particular nearest-neighbor solvation shell. As the appropriate Boltzmann weights are included in the Metropolis Monte Carlo sampling technique, the average value of the solvatochromic shift is obtained from a simple average over a chain E_i of size L of uncorrelated configurations, where $L = 40$ for all the systems considered here and E_i corresponds to the excitation energy obtained for the supermolecular configuration i.

2.2 Nearest-Neighbor Solvation Shells

The molecular structure of liquids are best analyzed using the concept of the radial distribution function (RDF). This is of particular importance in solute-solvent structures as it defines the solvation shells around the solute molecule[44, 47]. The RDFs represent fluctuations in the local density due to structure in the liquid. Specifically, the average density of atoms of type Y around atoms of type X is $\rho_{X-Y}(r) = \rho_Y G_{X-Y}(r)$, where $\rho_Y = (N_Y/V)$ is the density of type Y atoms, r is the $X - Y$ separation and $G_{X-Y}(r)$ is the RDF between atoms of type X and Y. For a solute-solvent system, the first atom of the RDF (type X) belongs to the solute molecule and the second atom (type Y) belongs to the solvent molecule. In the simulation, the $G_{X-Y}(r)$ is obtained by accumulating and normalizing histograms with the total number of atom pairs $X - Y$ found in a distance between $r - \delta r/2$ and $r + \delta r/2$,

$$G_{X-Y}(r) = \frac{HISTOGRAM[r - \frac{\delta r}{2}, r + \frac{\delta r}{2}]}{l n_X n_Y N_X (\frac{4\pi\rho_Y}{3})[(r + \frac{\delta r}{2})^3 - (r - \frac{\delta r}{2})^3]} \tag{2}$$

where l is the number of MC configurations analyzed during the simulation, n_X is the number of type X atoms in the solute, n_Y is the number of type Y atoms in the solvent, N_X is the number of solute molecules, N_Y is the number of solvent molecules and δr is the width of each bin of the histogram. If a liquid is structureless, then $G_{X-Y}(r) = 1$.

For a system consisting of one β-carotene in solution the type X of the RDF could be defined as both carbon (C) or hydrogen (H) atoms. Although

RDFs can be determined from diffraction experiments on liquid, such data are not currently available for β-carotene in solution. However, the calculated RDFs were used here to describe the distribution of the solvent molecules around the β-carotene and define its respective solvation shells. This structural analysis will be of great importance in the calculation of the absorption spectrum of β-carotene in solution. Quantum mechanical calculation of the absorption spectrum of a supermolecular system, composed of 901 molecules, over 40000 MC configurations is of course not possible. The alternative we suggested[27, 29] was to perform quantum mechanical calculations after the simulation, but using only a few selected number of solvent molecules[31, 38] and a selected number of MC configurations[28, 29, 30, 31, 32]. The number of solvent molecules included in the calculation is obtained from the analysis of RDFs, using all molecules surrounding the solute up to a certain solvation shell. This number is still large enough to preclude sophisticated *ab initio* calculations but lies well within the range of semiempirical methods. The number l of necessary MC configurations for ensemble average has already been reduced dramatically from 40000 to only 40, as discussed in the previous section. We shall show that this small number gives indeed converged result, and is a consequence of the markovian chain generated by MC simulations, as documented before[28, 29, 30, 32]. In figure 3, the two RDFs between C and H of β-carotene and the central C atom of acetone are shown as an example of the liquid structure obtained in the simulations. Irrespective of the solvent and its selected atom, all RDFs of β-carotene in solution, studied here, has the same shape with broad and low peaks (see figure 3). This just reflects the elongated geometry of the β-carotene and the wide spatial distribution of the carbon and hydrogen atoms. Therefore, these two RDF ($G_{C-Y}(r)$ and G_{H-Y}) can not help in the description of the distribution of solvation shells around the β-carotene.

The RDF between the center-of-mass of the solute and the solvent molecules, $G_{CM-CM}(r)$ is another natural possibility of describing the solvation shells around the β-carotene. In figure 4a, the RDF between the center-of-mass of β-carotene and acetone molecules is shown as an example of the liquid structure. This $G_{CM-CM}(r)$ presents a clear definition of four peaks that characterize the solvation shells around the center-of-mass of the β-carotene. The number of solvent molecules in each shell was obtained by integrating the peaks. In the case presented in figure 4a, 7 acetone molecules were found in the first shell (integrating until 6.35 $\mathring{A}$), 30 in the second shell (from 6.35 to 10.65 $\mathring{A}$), 46 in the third shell (from 10.65 to 13.85 $\mathring{A}$) and finally 108 in the fourth shell (from 13.85 to 18$\mathring{A}$). Figure 4b will be discussed soon below.

Figure 5 illustrates typical configurations, generated in the simulation, of one β-carotene surrounded by the 7 and 37 acetone molecules corresponding

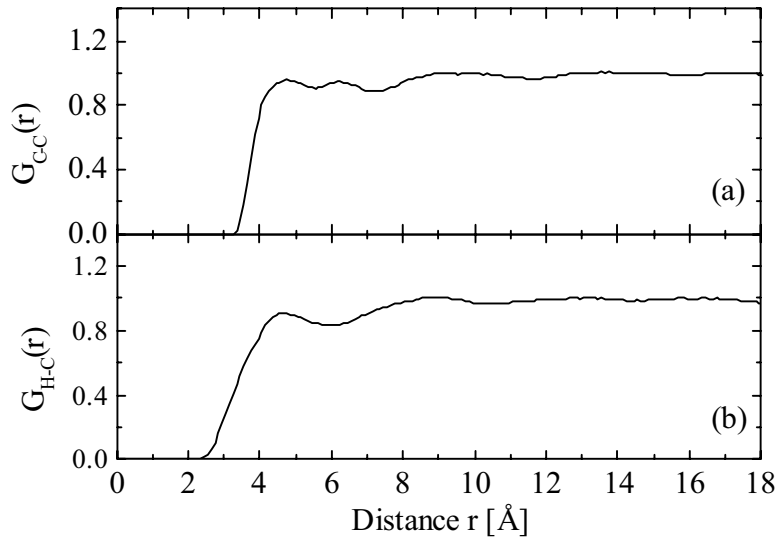

Figure 3: The calculated radial distribution function (RDF) between carbon atoms (a) and hydrogen atoms (b) of the β-carotene and carbon atoms of the acetone molecules, $G_{C-C}(r)$ and $G_{H-C}(r)$, respectively.

to the first and second peaks, respectively, defined by $G_{CM-CM}(r)$. Although the center-of-mass RDF presented peaks that *a priori* could be considered as solvation shells, certainly the figure 5 shows that, in the case of β-carotene as the solute, these peaks can not be considered as solvation shells around the solute. As expected, the solvent molecules were distributed only in the central part of the β-carotene, close to the center-of-mass. Of course, even considering the second or third solvation shells the $G_{CM-CM}(r)$ still gives a rather non-uniform distribution of solvent molecules around this elongated solute.

To analyze the solvation shells of elongated molecules in solution, it is necessary to define a different kind of RDF that does not grow in a spherical form, but consider the shape of the solute. Then, we suggest here a nearest-neighbor RDF between *all atoms* of the elongated solute and its nearest atom of each one of the N_Y solvent molecules, $G_{X-Nearest}(r)$. The $G_{X-Nearest}(r)$ was calculated using the same definition of equation 2, but changing the assignment of types X and Y. Now, all atoms of the solute molecule are assigned as an unique type X and after testing all atoms of the solvent molecule, the nearest atom of each solvent molecule is assigned as an unique type Y. Thus, a list of nearest-neighbors was built taking into account not a fixed atom or the center-of-mass of the solute, but all atoms in the solute molecule. With this new list of neighbors the shape of the elongate solute is taken into account in the distribution of solvent molecules. However for small solute, like, for instance, formaldehyde,

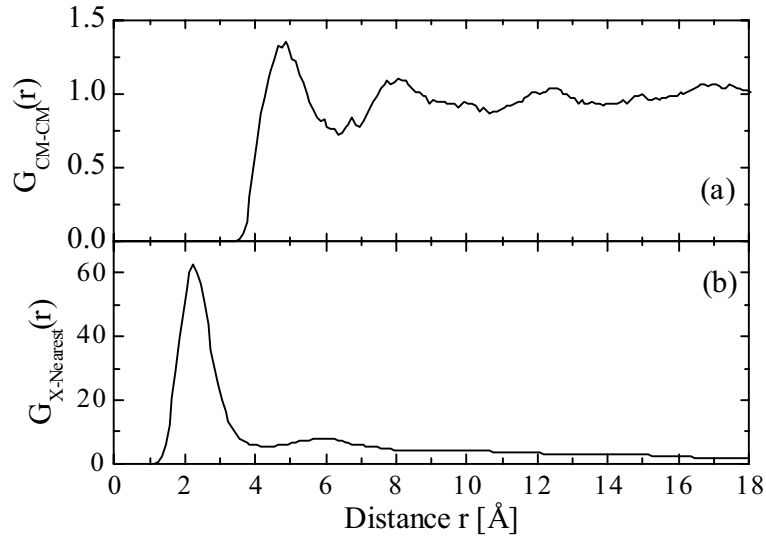

Figure 4: The calculated radial distribution function (RDF) between (a) the β-carotene center-of-mass and acetone center-of-mass, $G_{CM-CM}(r)$, and (b) all atoms of the β-carotene and its nearest atom of each acetone molecule, $G_{X-Nearest}(r)$.

this new list of neighbors generates a distribution of solvent molecules that is similar to that described by the center-of-mass distances. Thus, the nearest-neighbor distribution function generalizes the concept of solvation shells for a solute of any arbitrary shape. In figure 4b, the nearest-neighbor RDF of β-carotene in acetone is shown as an example of the liquid structure defined by $G_{X-Nearest}(r)$, that presents a clear definition of two solvation shells around the whole β-carotene. The number of solvent molecules in each solvation shell was obtained by integrating the peaks. In the case presented in figure 4b, 50 acetone molecules were found in the first solvation shell (integrating until 4.35 $\AA$) and 88 in the second shell (from 4.35 to 8.05 $\AA$). Figure 6 illustrates a typical configuration, generated in the simulation, of one β-carotene surrounded by 50 acetone molecules corresponding to the first solvation shell as defined by $G_{X-Nearest}(r)$. As expected, the solvent molecules are uniformly distributed around the β-carotene. For the other solvents, the minimum distance RDF presents the same shape as shown in figure 4b, a first well defined peak, a second less intense peak and a long structureless tail.

In table 3, a summary of the structural analysis is shown for all four solvents. Using the $G_{CM-CM}(r)$, the first neighborhood around the center-of-mass of the β-carotene is formed by either 5 isopentane or 7 acetone or 8 methanol or 9 acetonitrile. These solvent molecules are distributed approximately between 3.55 and 6.35 $\AA$ with maximum occurrency in the interval

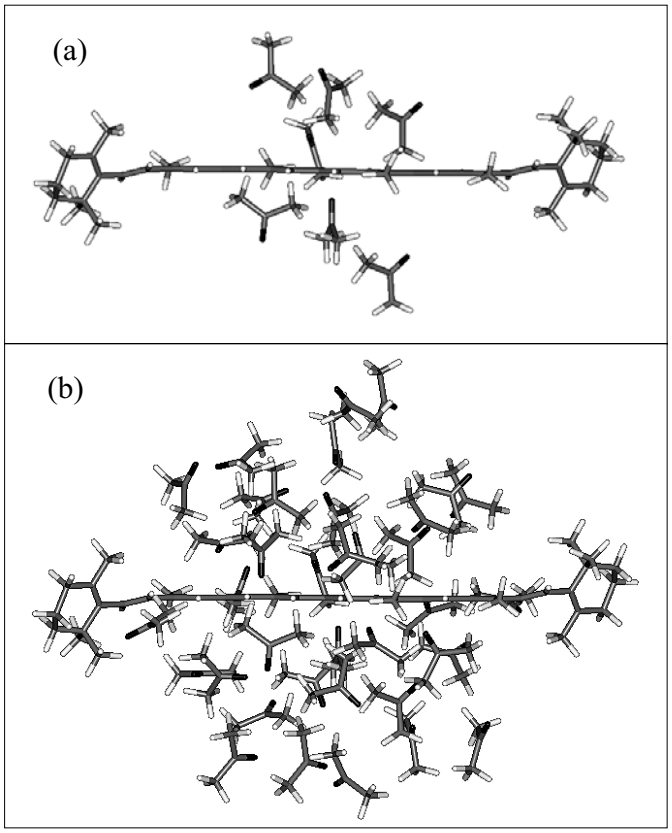

Figure 5: This illustration shows the β-carotene surrounded by (a) 7 acetone molecules and (b) 37 acetone molecules corresponding to the first and second peaks, respectively, defined by the center-of-mass RDF, $G_{CM-CM}(r)$.

between 4.35 $\mathring{A}$ (methanol) and 5.05 $\mathring{A}$ (isopentane). However, the first solvation shell around the entire β-carotene using the nearest-neighbor distribution is formed by 40 isopentane or 50 acetone or 69 methanol or 58 acetonitrile. These solvent molecules have their nearest atom distributed approximately between 1.35 and 4.45 $\mathring{A}$ with a large maximum occurrence in $\sim$ 2.30 $\mathring{A}$.

3 Quantum Mechanical Results

3.1 Solvatochromic Shifts

The quantum chemistry calculations are of the SCF type followed by configuration interaction over singly excited configuration state functions (CIS). This is the level of theory at which the INDO/S Hamiltonian was parametrized

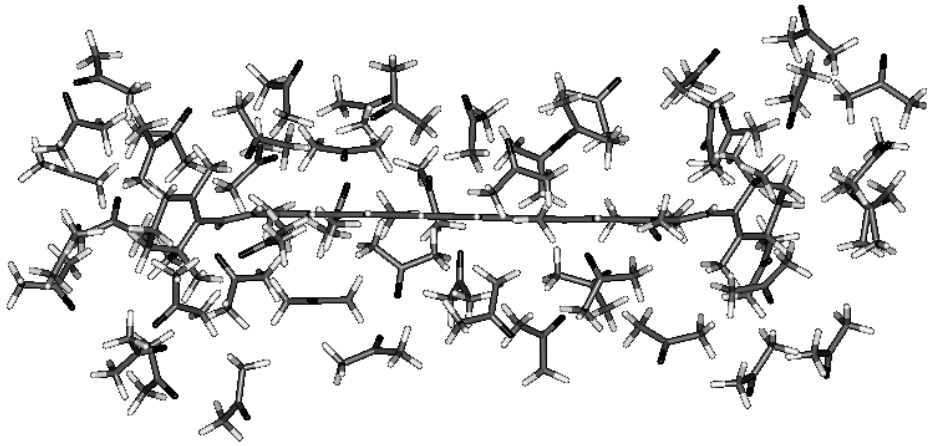

Figure 6: Typical supermolecule used in the QM calculations. This illustration shows the β-carotene surrounded by 50 first-neighbor acetone molecules corresponding to the first solvation shell as defined by the nearest-neighbor RDF, $G_{X-Nearest}(r)$.

Table 3
Structural information obtained from the first peak of the radial distribution functions. Distances are given in $\mathring{A}$ and N_s is the coordination number obtained from the integration of the first peak.

	$G_{CM-CM}(r)$				$G_{X-Nearest}(r)$			
Solvent	Start	Max.	End	N_s	Start	Max.	End	N_s
Isopentane	3.85	5.05	6.35	5	1.35	2.15	4.55	40
Acetone	3.55	4.65	6.35	7	1.35	2.25	4.35	50
Methanol	3.35	4.35	5.85	8	1.35	2.35	4.65	69
Acetonitrile	3.45	4.95	6.35	9	1.35	2.45	4.35	58

[60]. The low energy spectrum of β-carotene is dominated by excitations from the HOMO (π) molecular orbital to the LUMO (π^*) molecular orbital, where HOMO refers to the highest occupied molecular orbital and LUMO to the lowest unoccupied molecular orbital (illustrated in figure 7). As it can be seen the dipole moment is zero in both the ground and the excited state and the π orbitals are delocalized over the polyene chain. The gas phase absorption transition is not known experimentally. The value calculated here for the gas phase $\pi - \pi^*$ transition is 22230 cm^{-1}. In solvents of any polarity it is expected that this transition suffers a red shift. The magnitude of the shift, of course, depends on the solvent. To obtain these red shifts we have next performed the supermolecular calculations of β-carotene in acetone, acetonitrile, methanol and isopentane using the configurations generated by the MC simulation. As β-carotene is non-polar, it is expected that the solvation shift should not de-

pend on solvent molecules that are situated much beyond the first shell. Table 4, thus gives the calculated $\pi - \pi^*$ transition energies of β-carotene in the four solvents considered here. As it can be seen, in all solvents this transition suffers a red shift compared with the gas phase value. These results are in very good agreement with the experimental data. The relative positions are also correct for isopentane-methanol, for instance, but has a wrong sign (still within the statistical error) for the acetonitrile-acetone, if we use only the first shell of solvent molecules. Whereas the experimental result gives that the shift is larger for acetone over acetonitrile by 24 cm^{-1} the theoretical result as this stage gives the opposite by 12 cm^{-1}. Including half of the second shell changes the calculated value of acetone by only 3 cm^{-1}, showing that indeed this value is converged for the first shell. However, the other change for acetonitrile is slightly larger, by 36 cm^{-1}. As a result, including this half-shell corrects the

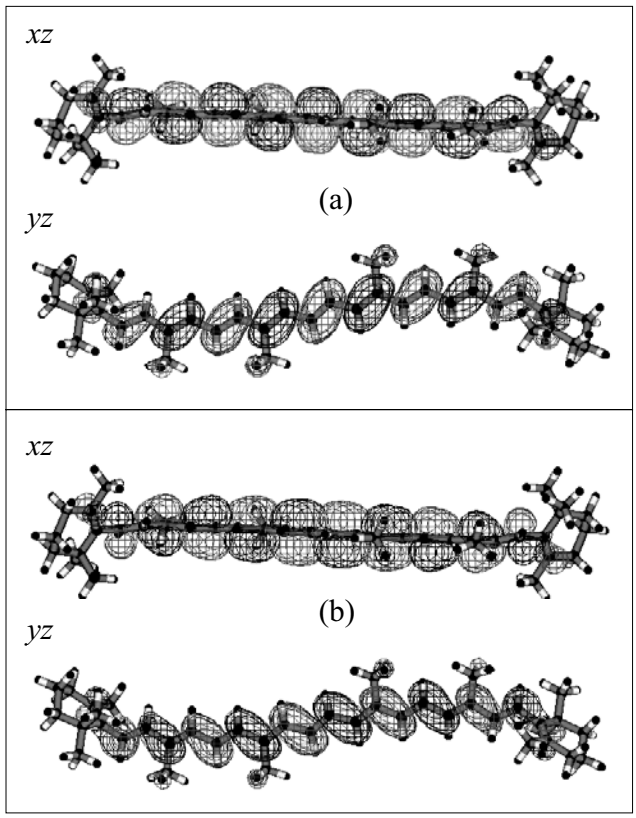

Figure 7: Illustration of the orbitals (a) HOMO and (b) LUMO involved in the first $\pi - \pi^*$ transition of β-carotene.

relative position of these two transitions. Note that the experimental shift of 24 cm^{-1} is now theoretically obtained as 21 cm^{-1}, in excellent agreement.

The results obtained for acetone, for instance, is a result of 40 QM INDO/CIS calculations of one β-carotene surrounded by 77 acetone molecules. Each QM calculation is thus a 2064-valence-electron problem. The largest calculation is for β-carotene in isopentane, that involves 2104 valence electrons. The qualitative relative shifts of the $\pi - \pi^*$ transitions are well reproduced in agreement with the experimental results. Note that the solvents are of miscellaneous type, involving both protic and non-protic, polar and non-polar and also of low and high polarity. In this direction, it should be noted that these shifts do not follow the increase in dielectric constants or, even, in polarity, as it can be checked from the results given in tables 1 and 4. Thus it is not clear that a description based on macroscopic parameters can be obtained. Abe and co-workers[42], have found an approximate correlation but only after excluding protic solvents. Similarly, the correlations are different for non-polar and polar solvents. In this paper, we explicitly calculate these values using a methodology that combines statistical mechanics and quantum mechanics. The consideration of dispersive interaction is of great importance. Analyzing the calculated results given in table 4, one may conclude that our approach is very successful in describing the relative shifts of this very challenging system. Before concluding, it is rather appropriate to discuss the convergence of the calculated result with respect to the statistics; i.e. with respect to the number of structures used in the QM calculations.

Table 4
Summary of the calculated and experimental results for the first $\pi - \pi^*$ absorption transitions of the β-carotene in gas phase and in solution.

Solvent	1st Shell		1st+half Shell		Experiment[42]
	N	$Transition$	N	$Transition$	
Vacuum	1	22230			
Acetone	$1+50$	22071 ± 17	$1+77$	22074 ± 17	22046
Acetonitrile	$1+58$	22059 ± 19	$1+92$	22095 ± 11	22070
Methanol	$1+69$	22143 ± 6	$1+90$	22143 ± 7	22247
Isopentane	$1+40$	22181 ± 4	$1+59$	22182 ± 4	22364

3.2 Statistical Convergence Analysis

In several previous applications we have shown that the auto-correlation function of the energy can be used to obtain statistically converged values from a small number of uncorrelated structures. In the present applications, only 40 QM calculations have been performed and it thus seems quite appropriate to discuss the convergence problem. Figure 8 shows the distribution of the

calculated individual transition energies for the case of β-carotene in acetone, including 50 acetone molecules (first shell). The calculated average value, as given before in table 4, is 22071 cm^{-1} and is shown as the horizontal line in Figure 8. This distribution clearly shows that a single structure can not describe the liquid situation. Although, some structures can give a transition energy that is close to the average, this is rather fortuitous and a few other structures give transition energies that are far from the average. It is necessary to consider the average of several calculations. Figure 9 shows how the average value approaches the convergence with increased number of structures used. As stated before, and now clearly seen in figure 8, the calculated $\pi - \pi^*$ transition energy is a converged value, after using only 35 configurations. The results obtained for all the other solvents are similar to those shown for the acetone case. Converged values are obtained and a single structure can not represent the statistical nature of the liquid. It may be worth mentioning that these results demonstrate that using gas-phase optimized geometries in solute-solvent situations is rather artificial and this single structure clearly can not represent the liquid environment. Using the statistically uncorrelated structures obtained from an analysis of the auto-correlation function of the energy we obtain converged results after only a few QM calculations. The spread of the calculated results can be used to obtain the contribution of the liquid structure to the line broadening[25, 28, 38]. Truly uncorrelated configurations are obtained only with an infinite separation, because the auto-correlation follows an exponential decay[27, 28, 29]. In most of our applications we have used structures that are less than 10% correlated. It has been discussed before that using more structures is important for decreasing the statistical error but has no effect on the converged average value[32]. Clearly, the same analysis can be used for the calculation of other properties[32].

4 Summary and Conclusions

The solvatochromic shifts of the $\pi - \pi^*$ transition of all-*trans*-β-carotene in different solvents have been studied using a sequential Monte Carlo/quantum mechanics (S-MC/QM) methodology. In this procedure we first generate the structures of the liquid using Metropolis MC simulation and perform the QM calculations in selected structures generated by the simulation. These structures are selected after an analysis of the relative statistical correlation between successive configurations. This leads to a large decrease of the number of structures used in the QM calculations, without affecting the average converged value. In the present application it is shown that including only 40 QM calculations gives statistically converged results. To deal with the very

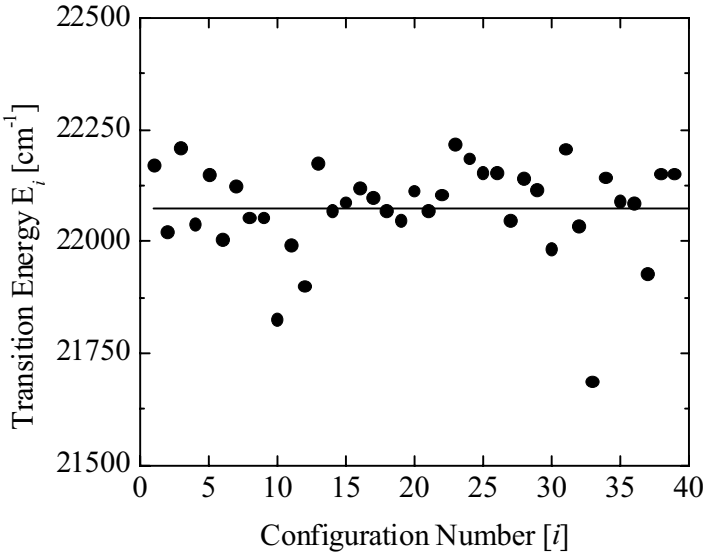

Figure 8: Distribution of the individual values of the $\pi - \pi^*$ transitions of the 40 MC configurations of β-carotene and 50 acetone molecules.

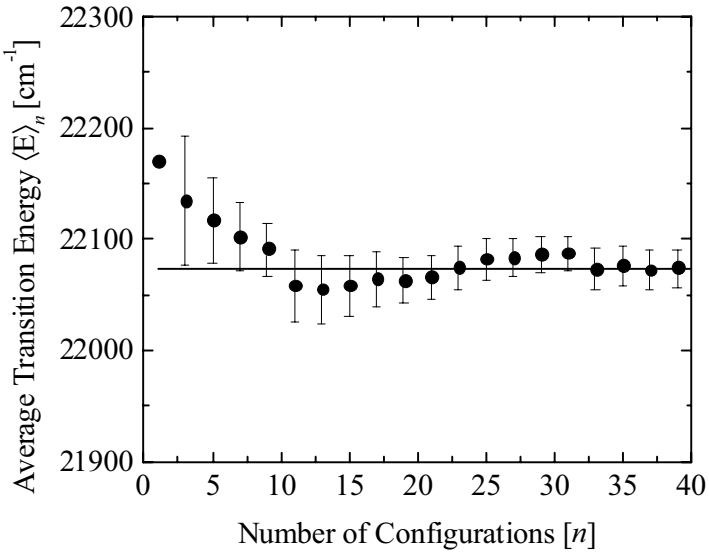

Figure 9: Convergence of the average value of the $\pi - \pi^*$ transition of β-carotene and 50 acetone molecules.

elongated shape of the all-*trans*-β-carotene solute molecule the MC simulation has been first extended to a large rectangular box. The use of a spherical radial distribution function is criticized in this case and we developed a nearest-neighbor distribution function between *all atoms* of the elongated solute and the nearest atom of each and every one of the solvent molecules. Although this has no effect on small and regular-shaped molecules it is of great importance in elongated solutes leading to a more appropriate distribution of neighbor molecules in solution. The nearest-neighbor distribution function, in fact, generalizes the concept of solvation shells for a solute of any arbitrary shape. Using only the first solvation shell the calculated results are found to be in very good agreement with the experimental results. However, to obtain the relative shifts in different solvents of varied properties, we found necessary to extend the number of solvent molecules. The relative shifts in isopentane, acetone, methanol and acetonitrile are calculated in excellent agreement with the experimental results. The different solvents are examples of systems of varied nature, differing in dielectric constants and covering a wide range of polarities, and including also polar and non-polar solvents.

As β-carotene itself is non-polar and the $\pi - \pi^*$ transition leads to a non-polar excited state, most of the solvatochromic shifts are consequence of the dispersive interaction. The solvation shift does not depend on solvent molecules that are situated much beyond the first solvation shell. In the present application, we find that inclusion of solvent molecules up to 6.0 $\mathring{A}$ is enough to give stable and accurate results, if the nearest-neighbor distribution function is used. This has also been found in the solvatochromic shifts of benzene in different solvents where the first solvation shell gives stable and accurate results[29]. This is, however, opposite to the case of formaldehyde (a polar molecule) in water (a protic solvent) where solvent molecules up to a distance of 10 $\mathring{A}$, were found to still affect the solvation shift[28]. The inclusion of dispersion interaction in the calculation of solvent effects has been recognized as one of the most important and difficult problems. It has been demonstrated[35] that although dispersion is a double excitation, calculation on a supermolecule that contains only single excitations includes dispersion interaction between the two subsystems when energy differences are taken between the ground state and low energy excited states in which single excitations dominate. Therefore the CIS calculations using supermolecular structures with explicit solute and solvent molecules seem to be an important step in this direction. Judging, from the qualitative and quantitative results of the solvatochromic shifts of β-carotene in different solvents, we are led to conclude that the most important contribution of dispersion is properly included.

5 Acknowledgments

It is our great privilege to dedicate this contribution to the memory of our teacher and friend Prof. Per-Olov Löwdin.This work is supported by CNPq, CAPES and FAPESP(Brazil).

References

[1] C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry*, Verlag Chemie, Weinheim, New York (1979).
C. Reichardt, *Chem. Rev.* **94,** 2319 (1994).

[2] C. J. Cramer and D. G. Truhlar, *Chem. Rev.* **99,** 2161 (1999).

[3] J. Tomasi and M. Persico, *Chem. Rev.* **94,** 2027 (1994).

[4] M. Orozco and F. J. Luque, *Chem. Rev.* **100**, 4187 (2000).

[5] L. Onsager, *J. Am. Chem. Soc.* **58,** 1486 (1936).

[6] J. G. Kirkwood, *J. Chem. Phys.* **2,** 351 (1934).
J. G. Kirkwood and F. H. Westheimer, *J. Chem. Phys.* **6,** 506 (1938).

[7] O. Tapia and O. Goscinski, *Molec. Phys.* **29,** 1653 (1975).

[8] J. L. Rivail and D. Rinaldi, *Chem. Phys.* **18,** 233 (1976).

[9] S. Miertus, E. Scrocco and J. Tomasi, *J. Chem. Phys.* **55,** 117 (1981).

[10] K. V. Mikkelsen, H Ågren and H. J. A. Jensen, *J. Chem. Phys.* **89,** 3086 (1988).

[11] E. Cances, B. Mennucci and J. Tomasi, *J. Chem. Phys.* **107,** 3032 (1997).

[12] A. Klamt and G. Schüürman, *J. Chem. Soc.Perkin Trans.* **2,** 799 (1993).
A. Klamt, V. Jonas, T. Burger and J. C. Lohrenz, *J. Phys. Chem. A* **102,** 5074 (1998).

[13] C. J. Cramer and D. G. Truhlar, *J. Am. Chem. Soc.* **113,** 8305 (1991).

[14] M. M. Karelson and M. C. Zerner, *J. Phys. Chem.* **96,** 6949 (1992).

[15] O. Christiansen and K. V. Mikkelsen, *J. Chem. Phys.* **110,** 1365 (1999).

[16] K. K. Baldridge and V. Jonas, *J. Chem. Phys.* **113,** 7511 (2000).
K. K. Baldridge, V. Jonas and A. D. Bain, *J. Chem. Phys.* **113,** 7519 (2000).

[17] J. Li, C. J. Cramer and D. G. Truhlar, *Int. J. Quantum Chem.* **77,** 264 (2000).

[18] A. Warshel and M. Levitt, *J. Mol. Biol.* **103,** 227 (1976)
A. Warshel, *J. Phys. Chem.* **83,** 1640 (1979).

[19] M.J. Field, P. A.Bash and M. Karplus, *J. Comput. Chem.* **11,** 700 (1990).

[20] J. Gao and X. Xia, *Science* **258,** 631 (1992).

[21] J. Gao, *J. Am. Chem. Soc.* **116,** 9324 (1994).

[22] J. Gao and K. Byun, *Theor. Chem. Acc.* **96,** 151 (1997).

[23] D. Bakowies and W. Thiel, *J. Phys. Chem.* **100,** 10580 (1996).

[24] M. A. Thompson, *J. Phys. Chem.* **100,** 14494 (1996).

[25] J. T. Blair, K. Krogh-Jespersen and R. M. Levy, *J. Am. Chem Soc.* **111,** 6948 (1989).

[26] J. Zeng, J. S. Craw, N. S. Hush and J. R. Reimers, *J. Chem. Phys.* **99,** 1482 (1993).

[27] K. Coutinho and S. Canuto, *Adv. Quantum Chem.* **28**, 90 (1997).

[28] S. Canuto and K. Coutinho, *Int. J. Quantum Chem.,* **77** 192 (2000).

[29] K. Coutinho, S. Canuto and M. C. Zerner, *J. Chem. Phys.* **112**, 9874 (2000).

[30] K. Coutinho, M. J. de Oliveira and S. Canuto, *Int. J. Quantum Chem.* **66**, 249 (1998).

[31] K. J. de Almeida, K. Coutinho, W. B. De Almeida, W. R. Rocha and S. Canuto, *Phys. Chem. Chem. Phys.* **3**, 1583 (2001).

[32] W. R. Rocha, K. Coutinho, W. B. De Almeida and S. Canuto, *Chem.Phys. Lett.* **335**, 127 (2001).

[33] P. Suppam and N. Ghoneim, *Solvatochromism*, Royal Soc. Chem., Cambridge (1997).

[34] N. Rösch and M. C. Zerner, *J. Phys. Chem.* **98**, 5817 (1994).

[35] S. Canuto, K. Coutinho and M. C. Zerner, *J. Chem. Phys.* **112**, 7293 (2000).

[36] W. Liptay, in *Modern Quantum Chemistry*, ed. O. Sinanoglu, Part II, p. 173, Academic Press, New York (1966).

[37] K. Coutinho, S. Canuto and M. C. Zerner, *Int. J. Quantum Chem.* **65**, 885 (1997).

[38] K. Coutinho and S. Canuto, *J. Chem. Phys.* **113**, 9132 (2000).

[39] K. Coutinho, N. Saavedra and S. Canuto, *J. Mol. Struct. (Theochem)* **466**, 69 (1999).

[40] K. J. de Almeida, W. R. Rocha, K. Coutinho and S. Canuto, *Chem.Phys. Lett.* **345**, 171 (2001).

[41] A. B. Myers and R. R. Birge, *J. Chem. Phys.* **73**, 5314 (1980).

[42] T. Abe, J.-L. Abboud, F. Belio, E. Bosch, J. I. Garcia, J. A. Mayoral, R. Notario, J. Ortega and M. Rosés, *J. Phys. Org. Chem* **11**, 193 (1998).

[43] C.-G. Zhan and D. M. Chipman, *J. Chem. Phys.* **109**, 10543 (1998), and refs. therein.

[44] P.A. Egelstaff, *An Introduction to the Liquid State*, Oxford Science Publ.,Oxford (1994).

[45] J. Applequist, *J. Phys. Chem.* **95**, 3539 (1991).

[46] M. C. Zerner, in *Problem Solving in Computational Molecular Science. Molecules in Different Environments*, Kluwer Acad. Publ., p. 249 (1997).

[47] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Oxford, (1987).

[48] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).

[49] D. R. Lide (ed.) *Handbook of Chemistry and Physics*, 73rd edition, 1992-1993, CRC-Press, Boca Raton, (1992).

[50] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *J. Am. Chem. Soc.* **118**, 11225 (1996).

[51] C. Sterling, *Acta Cryst.* **17,** 1224 (1964).

[52] A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
A. D. Becke, *J. Chem. Phys.* **98** 5648 (1993).

[53] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).

[54] W. L. Jorgensen, J. M. Briggs and M. L. Contreras, *J. Phys. Chem.* **94**, 1683 (1990).

[55] W. L. Jorgensen and J. M. Briggs, *Mol. Phys.* **63**, 547 (1988).

[56] W. J. Jorgensen, J. D. Madura and C. J. Swenson, *J. Am. Chem. Soc.* **106**, 6638 (1984).

[57] W. L. Jorgensen, J. D. Madura and C. J. Swenson, *J. Phys. Chem.* **90**, 1276 (1986).

[58] K. Coutinho and S. Canuto, *DICE* (version 2.8): A general Monte Carlo program for liquid simulation, University of São Paulo, (2000).

[59] M. C. Zerner, *ZINDO*: A Semi-empirical Program Package, University of Florida, Gainesville, FL 32611.

[60] J. Ridley and M. C. Zerner, *Theor. Chim. Acta* **32,** 111 (1973).