Study on the conformational equilibrium of the alanine dipeptide in water solution by using the averaged solvent electrostatic potential from molecular dynamics methodology

Francisco F. García-Prieto, Ignacio Fdez. Galván, Manuel A. Aguilar, and M. Elena Martín[a)]
Química Física, Edif. José María Viguera Lobo, Universidad de Extremadura, Avda. de Elvas s/n, 06071 Badajoz, Spain

(Received 14 June 2011; accepted 18 October 2011; published online 18 November 2011)

The ASEP/MD method has been employed for studying the solvent effect on the conformational equilibrium of the alanine dipeptide in water solution. MP2 and density functional theory (DFT) levels of theory were used and results were compared. While in gas phase cyclic structures showing intramolecular hydrogen bonds were found to be the most stable, the stability order is reversed in water solution. Intermolecular interaction with the solvent causes the predominance of extended structures as the stabilizing contacts dipeptide-water are favoured. Free-energy differences in solution were calculated and PPII, α_R, and C5 conformers were identified as the most stable at MP2 level. Experimental data from Raman and IR techniques show discrepancies about the relative abundance of α_R y C5, our results support the Raman data. The DFT level of theory agrees with MP2 in the location and stability of PPII and α_R forms but fails in the location of C5. MP2 results suggest the possibility of finding traces of C7eq conformer in water solution, in agreement with recent experiments. © *2011 American Institute of Physics*. [doi:10.1063/1.3658857]

I. INTRODUCTION

Small peptides have been historically employed as models in the study of the protein folding, being the understanding of this process a matter of utmost importance both experimentally and theoretically. Some of the most profusely studied peptides are the alanine dipeptide[1–36] (Ace-Ala-NMe, AD) and the simpler analogue 2-(formylamino)propanamide,[37] (ADA) where the terminal methyl groups have been substituted by hydrogen atoms. These peptides reproduce some of the most important structural features of the protein backbone. Being these minimal systems representative of the structure of proteins and being their natural medium aqueous solution, in vacuum results have a quite limited utility. Despite this, a large number of theoretical studies were carried out in absence of a surrounding medium.[1–6] The secondary structure of proteins is the result of the subtle interplay between intra- and intermolecular interactions and consequently the introduction of the interaction with the solvent is of critical importance. Some of the previous theoretical studies have tended to introduce this interaction by using a classical representation[7,8,33–35] or statistical mechanical integral equation theories[9,10] for the solvent. By using AD as a test system and by comparing its structures and stabilities in vacuum and in solution the effect of the surrounding medium can be better understood.

The stability of the different conformers of AD in gas phase has been the subject of several theoretical studies characterized by the use of different levels of theory and diverse and progressively more complete basis set. In the paper of Vargas *et al.*[11] a useful summary of the previous studies of the behaviour of this system in gas phase can be found. In agreement with experimental data,[38] the cyclic C7eq, C5, and C7ax conformers were found the most stable. It is widely accepted that the solvent effect, in a high-polarity solvent such as water, on the conformational equilibrium of AD is very pronounced giving rise to the destabilization of some structures present in gas phase and the modification of their dihedral ϕ and ψ angles to reach a more favourable disposition for the interaction with the solvent molecules. This reorientation of the CO and NH peptide bonds permits the formation of intermolecular hydrogen bonds with water molecules and stabilizes the polyproline-II (PPII) conformer.

In spite of the evident importance that the surrounding medium exerts on the structure of peptides, the variety of theoretical studies in solution is quite more limited than in gas phase and conclusions are not always coincident. Recent studies by Seabra *et al.*[18,19] and Kwac *et al.*,[20] and previously by Hu *et al.*,[21] compare the performance of different classical (MD) and semiempirical QM/MM methods in the determination of the conformational distribution of AD in water solution. Besides, dipolar coupling constant and numerically simulated amide I IR and VCD spectra were compared with experimental data. Their results show that QM/MM semiempirical methods do not always provide better results than sophisticate force fields, but equally it seems evident that the relative populations of the different conformers obtained by using classical MD strongly depend on the force field employed. This fact has been previously reported in the literature.[21–25] Nevertheless, probably due to the low level of calculation (purely classic or semiempiric), none of the

[a)]Author to whom correspondence should be addressed. Electronic mail: memartin@unex.es.

methods employed was able to accurately reproduce experimental data.[12–17] The multicanonical *ab initio* QM/MM study by Jono *et al.*[26] agreed with experimental results as it found four conformations present in aqueous solution (C5, PPII, C7eq, and α_R), but important discrepancies related with the distribution were reported. Once again, this discrepancy was probably due to the low *ab initio* level employed in the study (HF/3-21G). With a better accuracy in the AD (or ADA) description but without taking into account the microscopic solvent structure, PCM[27–30] and SCRF[28,31] calculations can be found in the literature. Probably due to the lack of consideration of specific solute-solvent interaction these methods fail in the prediction of the predominant species in aqueous solution even if the most significant changes with respect to the gas phase are achieved. In any case and in a general way, practically all the theoretical studies of AD (or ADA) in water solution agree in the remarkable variations that the potential energy surface shows when solvent effects are taken into account.

In our study, the interplay between the solute charge distribution and the solvent structure was modelled with the ASEP/MD method[39] (averaged solvent electrostatic potential from molecular dynamics). This method is especially adequate in the study of systems in which there exists an important correlation between solute polarization, solvent structure, and geometrical changes as it permits combining (1) a high-level *ab initio* description of the electronic structure of the solute, (2) the consideration of the mutual polarization of solute and solvent, (3) the location of minima and other stationary points on free-energy surfaces, (4) the calculation of free-energy differences between different solute-solvent geometries, and (5) the inclusion of specific interactions.

The rest of the paper is organized as follows: The method section covers a brief outline of the ASEP/MD methodology and the procedure followed in the calculation of free-energy differences. The Computational Details section presents the complete description of the carried out calculations. The results achieved are collected in the Results section. Finally, the last section brings together the conclusions reached throughout the paper.

II. METHOD

The ASEP/MD method[39] was applied in determining the AD electronic structure in aqueous solution. This method, developed in our laboratory, is a sequential QM/MM method where quantum mechanics (QM) and molecular mechanics (MM) techniques are alternated and not simultaneous. The key characteristic of the method is the use of the mean field approximation[40] that introduces the perturbation generated by the solvent into the solute molecular Hamiltonian in an averaged way. This perturbation is obtained from MD simulations in the form of the averaged electrostatic potential that the solvent generates on the solute. Once the perturbation is known the associated Schrödinger equation can be solved and one gets a new solute charge distribution to be used in the following MD. During the MD simulations the intramolecular geometry and charge distribution of all the molecules remain fixed. The MD simulations and QM calculations are iterated until the charge distribution of the solute and the solvent structure around it become mutually equilibrated. The ASEP/MD methodology has showed to be a powerful and efficient method that makes it possible to include the solvent influence on high-level quantum calculations and allows the accurate study of systems and processes where the correct treatment of the solute-solvent interactions is compulsory. Examples of these cases are UV-vis spectra,[41–44] reactivity[45] or conformational equilibrium.[46–48]

As usual in QM/MM methods the ASEP/MD Hamiltonian can be partitioned into the characteristic three terms:[49–55]

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM}, \tag{1}$$

corresponding to the quantum part $\hat{H}_{QM}$, the classical part $\hat{H}_{MM}$, and the interaction between them $\hat{H}_{QM/MM}$. In the current study, the quantum part includes only the dipeptide molecule and the classical part all the solvent molecules.

The energy and wave function of the solute molecule in solution are obtained by solving the effective Schrödinger equation:

$$(\hat{H}_{QM} + \hat{H}_{QM/MM})|\Psi\rangle = E|\Psi\rangle. \tag{2}$$

The interaction term, $\hat{H}_{QM/MM}$, takes the following form:

$$\hat{H}_{QM/MM} = \hat{H}^{elect}_{QM/MM} + \hat{H}^{vdw}_{QM/MM}, \tag{3}$$

$$\hat{H}^{elect}_{QM/MM} = \int dr \cdot \hat{\rho} \cdot \langle V_s(r;\rho)\rangle, \tag{4}$$

where $\hat{\rho}$ is the solute charge density operator, and the brackets denote a statistical average. The term $\langle V_s(r;\rho)\rangle$, named ASEP, is the averaged electrostatic potential generated by the solvent at the position r, and it is obtained from MD calculations in which the solute molecule is represented by a fixed geometry and the charge distribution ρ. Clearly, geometry and charge distribution could be different for successive MD and the process must be iterated until convergence is achieved. The term $\hat{H}^{vdw}_{QM/MM}$ is the Hamiltonian for the van der Waals interaction, supposed in many cases to have little effect on the solute wave function. For this reason, it is generally represented by a classical potential that depends only on the solute-solvent nuclear coordinates. In our case a Lennard-Jones potential was used. Only the electrostatic term enters into the electron Hamiltonian. Repulsion and dispersion contributions do not affect the solute electron wave function as, in our model, they depend only on the nuclear coordinates. Nevertheless they affect the geometry optimization and molecular dynamics.

For geometry optimization in solution, a technique described in a previous paper[56,57] was used, based on the use of the free-energy gradient method.[58–60] This technique has demonstrated its utility in the geometry optimization of ground and excited states of molecules in solution. At each step of the optimization procedure, the mean value of the total force, F, and the Hessian, H, averaged over a representative set of solvent configurations are calculated as the sum of the solute and solvent contributions and are used to obtain a new geometry by using the rational function optimization method.

The force and the Hessian read

$$F(r) = -\frac{\partial G(r)}{\partial r} = -\left\langle \frac{\partial E(r, X)}{\partial r} \right\rangle \approx -\frac{\partial \langle E(r, X) \rangle}{\partial r}, \quad (5)$$

$$H\left(r, r'\right) \approx \frac{\partial^2 \langle E(r, X) \rangle}{\partial r \partial r'}, \quad (6)$$

where $G(r)$ is the free energy; $E(r, X)$ is the potential energy sum of intra- and intermolecular (solute-solvent) contributions; and the brackets denote a statistical average over the solvent configurations, X. Technical details about the practical application of this method and its relation with other methodologies can be found in Refs. 56 and 57.

Once the minima are located in solution their relative stability must be determined. To this end the free-energy differences were calculated. Here, we follow a dual-level methodology, since the internal energy, geometry, and charges of the different conformers are obtained using QM methods but the solute-solvent interaction component of the free-energy differences is calculated classically. The standard free-energy difference between two conformers in solution is written as the sum of two terms,[44]

$$\Delta G = \Delta G_{\text{solute}} + \Delta G_{\text{int}} + \Delta V, \quad (7)$$

where ΔG_{int} is the difference in the solute-solvent interaction free energy between the two QM conformers and ΔG_{solute} is approximated as the *ab initio* energy difference between the two QM conformers (A and B) calculated using the gas phase solute molecular Hamiltonian, $\hat{H}_{\text{QM}}$, and the in solution wave functions:

$$\begin{aligned} \Delta G_{\text{solute}} &= E^B_{\text{QM}} - E^A_{\text{QM}} \\ &= \langle \Psi^B | \hat{H}_{\text{QM}} | \Psi^B \rangle - \langle \Psi^A | \hat{H}_{\text{QM}} | \Psi^A \rangle. \end{aligned} \quad (8)$$

The last term, ΔV, includes the difference in the zero-point energy (ZPE) and entropic contributions of the solute. This term is normally evaluated by applying the harmonic approximation to the vibrational modes of the solute in solution, and it needs the information provided by the Hessian matrix. Obtaining an accurate enough Hessian matrix may require large computational resources and we often decide to approximate the results by neglecting this term. It must be noted that ΔV refers only to the internal nuclear degrees of freedom of the solute. Thermal and entropic contributions to the solute-solvent interaction energy are included in ΔG_{int}.

The ΔG_{int} term was calculated using the free energy perturbation method.[61,62] The solute geometry was assumed to be rigid and a function of a perturbation parameter, λ, and the solvent was allowed to move freely. When $\lambda = 0$, the solute geometry and charges correspond to the initial state and when $\lambda = 1$ to the final state. For intermediate values, a linear interpolation is applied. For the alanine dipeptide system, a value of $\Delta\lambda = 0.025$ was used. That means that a total of 41 separate molecular dynamics simulations were carried out to determine the difference on the solute-solvent interaction free energy between each pair of conformers.

III. COMPUTATIONAL DETAILS

B3LYP/6-311+G*, B3LYP/aug-cc-pVDZ, and MP2/aug-cc-pVDZ calculations have been performed in order to study the conformational equilibrium of the AD molecule in gas phase and in water solution. In solution calculations were carried out with the ASEP/MD method. As it was exposed before, ASEP/MD allows, through the combination of molecular dynamics simulations and *ab initio* quantum calculations, a correct description of the solute-solvent interactions and of the effect that the solvent causes on the structural features of the solute. It is worth remembering that this result can be reached with a reduced number of high-level quantum calculations, overcoming the main drawback of conventional QM/MM methods. *Ab initio* calculations were carried out with the GAUSSIAN 98 package of programs[63] both in vacuum and in solution and Moldy[64] was employed for the molecular dynamics simulations. The simulations contain one AD molecule and 300 water molecules at fixed intramolecular geometry in a cubic box of 21 Å side. Given the number of molecular dynamics simulations that the ASEP/MD procedure implies (15-20 for the ASEP/MD iterative process plus the molecular dynamics needed for the calculation of ΔG_{int}), the dimension of the box results from a compromise between the sufficient number of solvent molecules in order to conveniently consider the solute environment and the computational cost. The Lennard-Jones parameters for the solute molecule were taken from the OPLS-AA (optimized potentials for liquids simulations, all atoms) force field[65] and solute atomic charges were calculated using the CHELPG method.[66,67] For water molecules, the TIP3P model[68] was used. Periodic boundary conditions were applied and a spherical cutoff was used to truncate the solute-solvent and solvent-solvent interactions at 9.7 Å. The electrostatic interaction was calculated with the Ewald method and the temperature was fixed at 298 K with the Nosé-Hoover[69] thermostat. Each simulation was run for 75 ps with a time step of 0.5 fs. From the 150 000 steps, 50 000 were used for equilibration and the final 100 000 for production. In solution final results were obtained by averaging the last five ASEP/MD cycles, and therefore they represent a 250 ps average.

The atom numbering and representative dihedral angles of the AD molecule are displayed in Figure 1. In the structure

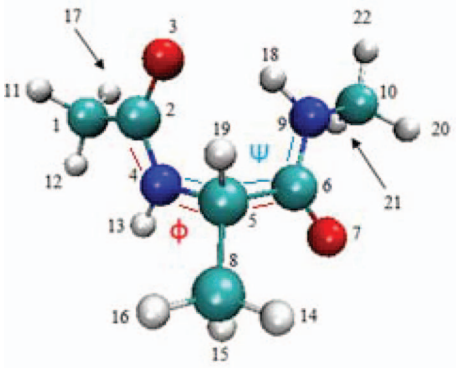

FIG. 1. Atom numbering and representative dihedral angles for the alanine dipeptide.

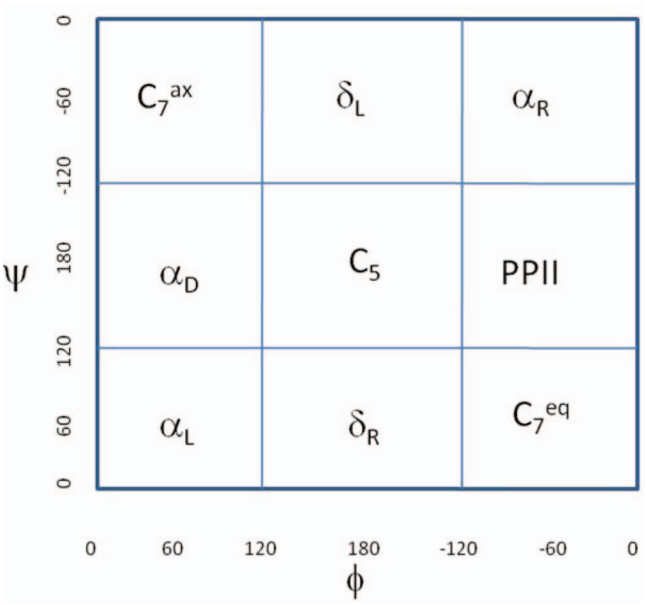

FIG. 2. One of the possible nomenclature systems for the different regions of the Ramachandran map in terms of the dihedral angles ϕ and ψ.

of peptides and proteins the backbone disposition is referred according to the values of the dihedral angles ϕ and ψ (C2N4C5C6 and N4C5C6N9, respectively). Nine different regions can be defined in the so-called Ramachandran map in terms of these angles and their nomenclature gives the label for the different backbone structures. These regions are named with the Greek alphabet from alpha to epsilon letters and their approximate values are displayed in Figure 2. Nevertheless some of the most representative peptide structures have specific names trying to describe their most evident structural characteristic. So, β and γ are labelled as C5 and C7 making reference to the five- and seven-membered rings present in the structures. Besides, *ax* or *eq* subscripts are added to C7 depending on the axial or equatorial position of the methyl group with respect to the ring.

IV. RESULTS AND DISCUSSION

A. Gas phase

As in previous MP2 and discrete Fourier transform (DFT) studies, in gas phase, six stable structures have been located at the three levels of calculation employed in this paper.[71] Their dihedral angles and their relative energies are collected in Table I. DFT and MP2 yield equivalent values of ϕ and ψ for all the minima, with the exception of the δ_R conformer. The MP2/aug-cc-pVDZ calculation gives for δ_R dihedral angles shifted from their characteristic values and actually corresponding to an α_R disposition (we will refer to this minimum as $\delta_R(\alpha_R)$). This fact has been previously referred to in the literature and our MP2 results completely agree both structurally and energetically with those published by Vargas *et al.*[11] at the same level of calculation. DFT calculations do not show this behaviour with neither of the employed basis sets and δ_R presents values around -115° and 14° for ϕ and ψ, respectively. It seems[70] that the energy surface around the δ_R minimum is quite flat and its location critically depends on the level of calculation. In our study, the discrepancy between MP2 and B3LYP must be attributed to the different level of theory, as DFT results are similar with both basis sets and different from MP2. We will return to this point in the discussion of in-solution results.

From an energetic point of view, the stability order is basically the same independently of the level of calculation. It is only to be noted the energetic degeneration showed by C7ax and δ_R at B3LYP/6-311+G* level of calculation. This fact is not found at B3LYP/aug-cc-pVDZ and given that the geometries obtained with both sets are equivalent, the discrepancy must be attributed to the 6-311+G* basis set. In a general way, structures with intramolecular hydrogen bonds are the most stable and in this way C7eq, C5, and C7ax present lower energies than δ_R, α_L, and δ_L. Our results are in agreement with experimental[38] studies at room temperature in CCl_4 solution and at low temperature in Ar matrices where C7eq and C5 are the predominant species. This behaviour is characteristic of the gas phase and important changes in the stability order are expected when the interaction with the solvent is considered. Intramolecular hydrogen bonds show oxygen-hydrogen distances equivalent at the three levels of calculation.[71]

B. Water solution

In water solution the landscape is radically different. Experimental results establish that the predominant species are extended structures as these favour mayor interaction with a polar and protic solvent such as water. In fact, a new structure, called PPII, turns out to be the more abundant, followed by α_R and to some extent by C5. A recent study by Takekiyo

TABLE I. Dihedral angles (in degrees) and relative energies (in kcal/mol) for the different conformers of the alanine dipeptide in gas phase.

	B3LYP/6-311+G*			B3LYP/aug-cc-pVDZ			MP2/aug-cc-pVDZ		
	ϕ	ψ	ΔE	ϕ	ψ	ΔE	ϕ	ψ	ΔE
C7eq	− 83.3	76.2	0.00	− 82.8	73.5	0.00	− 82.5	76.3	0.00
C5	− 154.8	158.7	1.07	− 155.0	158.3	0.99	− 161.1	157.1	1.70
C7ax	74.0	− 57.4	2.63	72.6	− 55.0	2.27	73.8	− 53.7	2.28
δ_R	− 115.8	13.9	2.63	− 116.0	15.3	2.83	− 82.8	− 9.6	3.09
α_L	72.4	19.1	5.41	71.7	18.9	5.52	64.1	30.1	4.44
δ_L	− 165.1	− 44.2	6.46	− 163.4	− 44.3	6.49	− 164.6	− 38.2	6.42

TABLE II. Dihedral angles (in degrees) and relative free energies (in kcal/mol) for the different conformers of the alanine dipeptide in water solution.

	B3LYP/6-311+G*			B3LYP/aug-cc-pVDZ			MP2/aug-cc-pVDZ		
	ϕ	ψ	ΔG	ϕ	ψ	ΔG	ϕ	ψ	ΔG
α_L	57.9	39.7	− 0.30	56.4	40.9	0.45	55.2	38.1	0.40
C7ax	70.7	− 57.0	0.00	70.1	− 53.9	0.00	70.7	− 54.9	0.00
δ_L	− 157.7	− 59.7	− 1.62	− 155.7	− 61.8	− 0.90	− 166.8	− 47.4	1.04
C7eq	–	–	–	–	–	–	− 81.9	85.9	0.44
C5	–	–	–	–	–	–	− 153.4	154.7	− 2.62
α_R	− 76.8	− 18.9	− 4.77	− 76.3	− 19.0	− 3.61	− 72.6	− 18.0	− 2.75
PPII	− 63.9	150.9	− 5.80	− 63.3	150.7	− 4.33	− 59.3	151.7	− 3.98

et al.[16] proposes as well the presence of the C7eq conformer both in aqueous solution and in chloroform. In any case, all of the experimental studies in water point out the PPII conformer as the predominant. PPII presents characteristic dihedral angles of around -60° and 150° for ϕ and ψ, respectively, and the most favourable interactions with solvent molecules. This situation contrasts with the trend in gas phase where the most stable conformers are cyclic structures thanks to the presence of intramolecular hydrogen bonds. Our three levels of calculation agree in the location of the PPII structure in water solution and in its assignment as the most stable form followed by the α_R species. Dihedral angles and free-energy differences between the minima located in water solution in this paper are collected in Tables II and III. As it has already been indicated, DFT and MP2 level of theory were considered and two basis sets were employed. For the lower accuracy level, B3LYP/6-311+G*, two sets of calculations were performed including or neglecting the ΔV contribution. The largest value of this term is 0.26 kcal/mol and as can be observed in Table III this component does not introduce variations in the stability order of the different AD conformers. For this reason and given the additional computational cost needed for its evaluation, we neglect the ΔV term in the rest of calculations.

It is interesting to note that MP2/aug-cc-pVDZ calculations identify, apart from the new PPII structure, the same minima in solution as in gas phase. In total, seven solvated species can be located.[65] Besides, the $\delta_R(\alpha_R)$ conformer in gas phase is consolidated in solution as a typical α_R. Regarding the DFT calculations, independently of the basis set employed, it is found that C7eq and C5 structures suffer evolution towards PPII, and finally only five conformers in solution are located. In this way, DFT results fail to identify species similar to C5 and C7eq although they coincide with MP2 in the location and stability of the two most representative conformers in solution. The typical δ_R conformer located in gas phase modifies its structure in water and finally adopts dihedral angles corresponding to α_R, has happened with MP2.

In sum, DFT and MP2 agree in the assignment of PPII and α_R conformers as the two most representative forms in water solution. The α_R conformer appears in aqueous solution from the structural modification of δ_R; DFT fails in the location of C5 and C7eq minima in solution as they evolve towards PPII. MP2/aug-cc-pVDZ is the only level that identifies the C5 conformer as the third species in stability in agreement with experimental data.

Dihedral angles (ϕ and ψ) are equivalent at the three levels of calculation for the structures subject of comparison. DFT dihedral angles are more similar between them and slightly different with respect to the MP2 results and consequently the differences must be attributed to the level of theory and not to the employed basis set. Turning now to energetic aspects, the stability of the different conformers in solution is the result of the interplay between two components (see Eq. (7)): the internal energy component (ΔG_{solute}) and the solvation energy (ΔG_{int}). In turn, the magnitude of the internal energy depends on the geometrical change in the solute and on the polarization of its wave function. The analysis of Tables I and III highlights that the internal energy in solution and the energy in gas phase follow a similar trend (only the exchange between C5 and C7ax can be noted). Consequently, the stability differences found in water solution and gas phase are due to the contribution of the solvation energy. Besides,

TABLE III. Free-energy components (in kcal/mol) for the different conformers of the alanine dipeptide in water solution. ΔG^* value includes the ΔV component.

	B3LYP/6-311+G*				B3LYP/aug-cc-pVDZ			MP2/aug-cc-pVDZ		
	ΔG_{solute}	ΔG_{int}	ΔG	ΔG^*	ΔG_{solute}	ΔG_{int}	ΔG	ΔG_{solute}	ΔG_{int}	ΔG
α_L	9.61	− 9.91	− 0.30	− 0.56	9.43	− 8.98	0.45	8.32	− 7.91	0.40
C7ax	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
δ_L	11.37	− 12.98	− 1.62	− 1.74	11.56	− 12.46	− 0.90	12.51	− 11.46	1.04
C7eq	–	–	–	–	–	–	–	− 1.32	1.75	0.44
C5	–	–	–	–	–	–	–	1.24	− 3.87	− 2.62
α_R	7.84	− 12.62	− 4.77	− 4.77	9.18	− 12.79	− 3.61	10.39	− 13.15	− 2.75
PPII	8.28	− 14.07	− 5.80	− 5.81	9.26	− 13.59	− 4.33	9.68	− 13.66	− 3.98

TABLE IV. Calculated population (in %) for the different conformers of the alanine dipeptide in water solution. Experimental Raman and IR populations are showed as well.

	MP2/aug-cc-pVDZ	B3LYP/aug-cc-pVDZ	B3LYP/6-311+G*	IR[a]	Raman[a]
PPII	81.5	77.0	85.0	60.0	76.0
α_R	10.2	22.7	15.8	11.0	18.0
C5	8.14	–	–	29.0	6.0
C7eq	0.05	–	–	–	–
Others	0.11	0.3	0.2	–	–

[a]Reference 12.

it is interesting to note that there exists a negative correlation between the internal energy and the solvation component, that is, the less stable the internal structure of the conformer, the greater the solvation energy. In this sense, the PPII structure exhibits the most negative solvation energy as it favours interaction with solvent molecules, and one of the largest values of internal energy as corresponds to extended structures and in the same trend as in gas phase.

Even if all the three sets of calculation agree in the identification of the two more representative conformers present in water solution, there exist some differences in the stability order of the less stable species. Nevertheless, as the energy differences between them are minimal and they are not experimentally identified, further discussion would be little conclusive.

The conformational distribution of AD in water solution obtained from the calculated free energy values is showed in Table IV along with IR and Raman[12] experimental data. The calculated abundance of the low-stability conformers is summed up for all of them and collected under the epigraph *others*. As it can be observed the population of these species is practically negligible and the three predominant conformers are PPII, α_R, and C5 in agreement with experimental data. Nevertheless some differences can be observed with respect to the population of these conformers. Raman and IR experiments agree in identifying PPII as the predominant conformer but they differ in the relative populations of C5 and α_R. The trend of our calculation is a slight overestimation of the PPII population especially marked in the case of DFT/6-311+G* level of calculation. At DFT/aug-cc-pVDZ level the PPII population fits better with experimental data thanks to a larger abundance of α_R. As it was previously discussed, it is necessary to go to MP2 level to find C5, and to some extent C7eq. With respect to the abundance order, our results support the Raman experimental data as the populations follow the order: PPII $>$ α_R $>$ C5 even if PPII and C5 populations seem to be slightly overestimated. In order to check the convergence of the results, some test runs were performed with 900 water molecules. The stability order of the different conformers is maintained even though the C5 conformer is slightly destabilized with respect to PPII and α_R. It can also be noted that, in agreement with Takekiyo *et al.*,[16] the seven-membered ring conformer, C7eq, could be present in water solution even if its abundance seems to be minimal.

It can be interesting to compare our MP2 results with a previous PCM study,[29] where a similar level of calculation was used (MP2/cc-pVTZ//MP2/6-31G**). This study predicts that the C5 conformer is the predominant form in water solution followed by PPII and α_R. The difference in the relative stability order between our results and PCM evidences the importance of taking into account the specific solute-solvent interactions.

In Figure 3 are collected the radial distribution functions for C5, at MP2 level of calculation, corresponding to the interactions O3-Hw, O7-Hw, H13-Ow, and H18-Ow. Even though the intramolecular hydrogen bond O7-H13 is longer in water solution than in gas phase (2.35 Å vs. 2.23 Å, see Tables I and II in supplementary material), the existence of this hydrogen bond affects strongly the interaction of H13 with the solvent. This fact can be observed by comparing the rdfs of H13 and H18. The first peak, corresponding to the first solvation shell, is notably lower for H13 than for H18 and its position is slightly shifted to longer distances, from 1.7 Å to 1.9 Å.

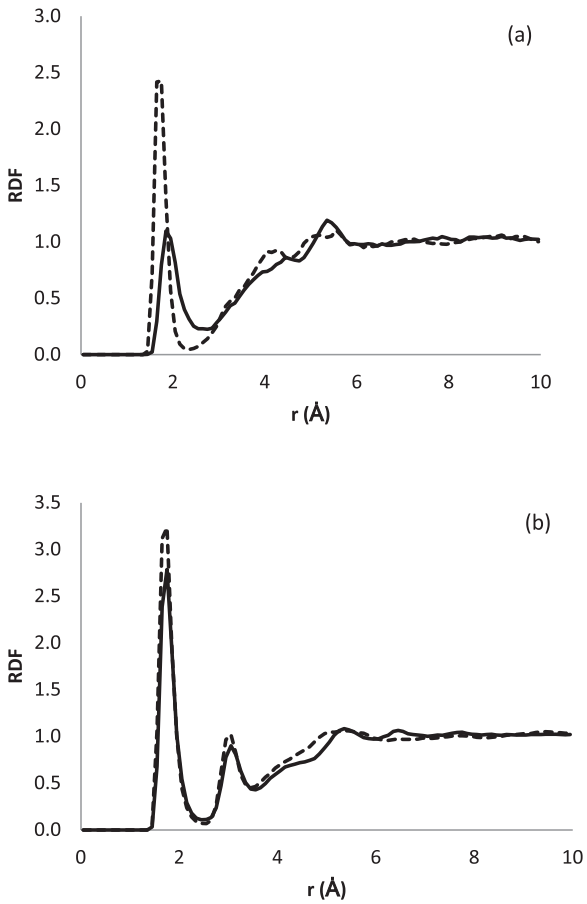

FIG. 3. Radial distribution function for the C5 conformer of the alanine dipeptide: (a) full line H13-Ow, dashed line H18-Ow and (b) full line O7-Hw, dashed line Hw-O3.

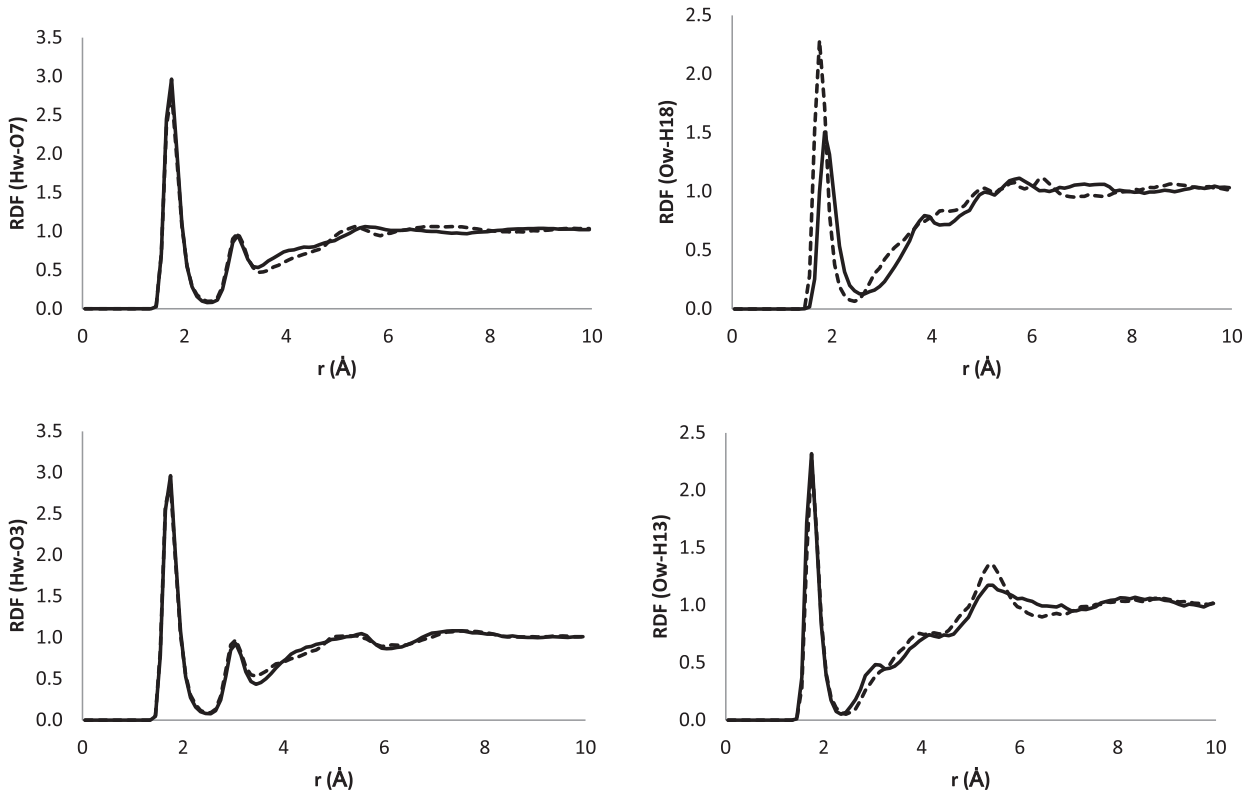

FIG. 4. Radial distribution functions for PPII (dashed line) and α_R (full line) conformers of the alanine dipeptide.

All these facts indicate a more loose solvent structure around H13 as a consequence of its implication in the intramolecular hydrogen bond. On the contrary, the similar shape showed for the rdfs O7-Hw and O3-Hw indicates a smaller modification of the O7-water interaction in spite of its involvement in the O7-H13 hydrogen bond. In any case, the first peak height and coordination number are smaller for rdf(O7-Hw) (2.78 Å and 2.7) than for rdf(O3-Hw) (3.23 Å and 3.0). Despite the cyclic structure of C5, the presence of three clear hydrogen bonds with water molecules gives stability in solution. In summary, the C5 conformer is located in water solution at MP2, and not at DFT level of theory due to a combination of two conditions. First, the better treatment of the N-H–O interaction showed by MP2 with respect to DFT (we will refer again to this behaviour) and second, the formation of three intermolecular hydrogen bonds with the solvent as it gives stability in solution.

It is noteworthy the different behaviour showed by MP2 and DFT with respect to the δ_R conformer in gas phase and in solution. If in gas phase MP2 and DFT, independently of the basis set employed, yield different geometries for δ_R, in water solution, on the contrary, all of them coincide in the location of a clear α_R structure. In gas phase, MP2 results provide an intramolecular N9-H18–O3 hydrogen bond distance of 2.96 Å (angle of 108.43°) in contrast with a distance of 3.43 Å (angle of 118.16°) at DFT level. It seems that MP2 favours more than DFT the formation of this kind of interaction and forces the rotation of ϕ and ψ dihedral angles reaching an α_R structure whereas DFT optimisation remains in δ_R values. This is the second evidence, in our study, about the different treatment of the N-H–O interaction showed by MP2 and DFT theories. In solution, an α_R structure is obtained at both levels of calculation as it permits a better interaction with water molecules.

Figure 4 collects some rdfs corresponding to the interaction of the PPII and α_R conformers with water molecules. The rdfs are completely equivalent at the three levels of calculation (MP2/aug-cc-pVDZ, B3LYP/6-311+G*, and B3LYP/aug-cc-pVDZ) and only those corresponding to the B3LYP/aug-cc-pVDZ calculation are displayed. It can be observed that all the graphs are similar except for the pair H18-Ow. For PPII a clear hydrogen bond between H18 and the water oxygen is observed at 1.75 Å whereas this interaction is not so important for the α_R conformer: its rdf first peak appears shifted to 1.85 Å and the height is notably lower. Although the coordination numbers for both conformers are equivalent, circa 1.0, the shape of their radial distribution functions for the pair H18-Ow indicates a more structured solvent and a stronger solvent-solute interaction for PPII. In conclusion, the PPII conformer shows four clear hydrogen bonds with the water solvent in contrast to the less structured solvent around α_R. This larger interaction of PPII with the solvent seems to be the reason why of the two extended structures, PPII is the predominant conformer in water solution. The larger values of the interaction free energy showed by PPII with respect to α_R in Table III agree with the previous statement.

Regarding the geometrical aspect, the comparison between the geometric parameters for the different conformers of AD, in gas phase and in solution, permits us to analyze the solvent influence on the AD geometry. The most significant

changes are related with the distances of the peptide bonds where C-O distances increase in solution whereas C-N distances decrease.[71] This behaviour, beside the slight increase of the peptide bond planarity, is compatible with the increase in stability of the charge-separated form of the peptide bond in solution. These results agree with the recently published results about the stability of the Cys-Asn-Ser tripeptide[48] in solution and other published studies.[29]

V. CONCLUSIONS

The structure and stability of the different alanine dipeptide conformers have been studied in gas phase and in water solution by using the ASEP/MD method. Three levels of calculation were chosen, B3LYP/6-311+G*, B3LYP/aug-cc-pVDZ, and MP2/aug-cc-pVDZ. In gas phase, the results are mostly independent of the level of theory and basis set employed and six conformers are located. Stability and geometry are found equivalent for the three sets of calculations except for the δ_R conformer which structure at MP2 level of calculation has dihedral angles resembling α_R. Cyclic conformers are identified as the more stable, being the stability order: C7eq > C5 > C7ax > δ_R > α_L > δ_L. The situation is radically different in water solution. The trend is a destabilization of cyclic structures in favour of the extended forms as they permit a better interaction with the solvent molecules. In this sense, an extended structure, called PPII, is experimentally identified as the most stable and both MP2 and B3LYP calculations agree in the assignation of this conformer as the principal form in water solution. With a lower population, α_R and C5 conformers are also present in solution. Our MP2 results support Raman data identifying α_R and C5 as the second and third more abundant conformers in water. It is worth noting that the B3LYP level fails in the location of C5 in solution as it evolves towards a PPII-like structure. Consequently, only the MP2 level of calculation is able to locate all the experimentally identified conformers in water solution. On the basis of our MP2 results and in agreement with Takekiyo *et al.*,[16] the C7eq conformer can be present in some extent even if its population according to our calculations would be minimal.

ACKNOWLEDGMENTS

This work was supported by the CTQ2008-06224/BQU Project from the Ministerio de Educación y Ciencia of Spain, co-financed by the European Regional Development Fund (ERDF). The authors acknowledge the financial support from the Consejería de Economía, Comercio e Innovación of the Junta de Extremadura. I.F.G. acknowledges the European Social Fund as well. F.F.G.-P. acknowledges a fellowship from the Ministerio de Educación y Ciencia.

[1] J. N. Scarsdale, C. Can Alsenoy, V. J. Klimkowski, L. Schäfer, and F. A. Momany, J. Am. Chem. Soc. **105**, 3438 (1983).
[2] H.-J. Böhm and S. Brode, J. Am. Chem. Soc. **113**, 7129 (1991).
[3] I. R. Gould and P. A. Kollman, J. Phys. Chem. **96**, 9255 (1992).
[4] K. J. Jalkanen and S. Suhai, Chem. Phys. **208**, 81 (1996).
[5] M. D. Beachy, D. Chasman, R. B. Murphy, T. A. Halgren, and R. A. Friesner, J. Am. Chem. Soc. **119**, 5908 (1997).
[6] D. M. Philipp and R. A. Friesner, J. Comput. Chem. **20**, 1468 (1999).
[7] M. Mezei, P. K. Mehrotra, and D. L. Beveridge, J. Am. Chem. Soc. **107**, 2239 (1985).
[8] A. Anderson and J. Hermans, Proteins **3**, 262 (1988).
[9] B. M. Pettitt and M. Karplus, Chem. Phys. Lett. **12Z**, 194 (1985).
[10] B. M. Pettitt and M. Karplus, J. Phys. Chem. **92**, 3994 (1988).
[11] R. Vargas, J. Garza, B. P. Hay, and D. A. Dixon, J. Phys. Chem. A **106**, 3213 (2002).
[12] J. Grdadolnik, S. G. Grdadolnik, and F. Avbelj, J. Phys. Chem. B **112**, 2712 (2008).
[13] V. Madison and K. D. Kopple, J. Am. Chem. Soc. **102**, 4855 (1980).
[14] M. A. Mehta, E. A. Fry, M. T. Eddy, M. T. Dedeo, A. E. Anagnost, and J. R. Long, J. Phys. Chem. B **108**, 2777 (2004).
[15] C.-D. Poon, E. T. Samulski, C. F. Weise, and J. C. Weisshaar, J. Am. Chem. Soc. **122**, 5642 (2000).
[16] T. Takekiyo, T. Imai, M. Kato, and Y. Taniguchi, Biopolymers **73**, 283 (2004).
[17] C. F. Weise and J. C. Weisshaar, J. Phys. Chem. B **107**, 3265 (2003).
[18] G. M. Seabra, R. C. Walker, and A. E. Roitberg, in *Solvation Effects on Molecules and Biomolecules*, edited by S. Canuto (Springer Science, Berlin, 2008), p. 507.
[19] G. M. Seabra, R. C. Walker, and A. E. Roitberg, J. Phys. Chem. A **113**, 11938 (2009).
[20] K. Kwac, K.-K. Lee, J. B. Han, K.-I. Oh, and M. Cho, J. Chem. Phys. **128**, 105106 (2008).
[21] H. Hu, M. Eltsner, and J. Hermans, Proteins **50**, 451 (2003).
[22] Y. Mu and G. Stock, J. Phys. Chem. B **106**, 5294 (2002).
[23] Y. Mu, D. S. Kosov, and G. Stock, J. Phys. Chem. B **107**, 5064 (2003).
[24] A. N. Drozdov, A. Grossfield, and R. V. Pappu, J. Am. Chem. Soc. **126**, 2574 (2004).
[25] P. E. Smith, J. Chem. Phys. **111**, 5568 (1999).
[26] R. Jono, Y. Watanabe, K. Shimizu, and T. Terada, J. Comput. Chem. **31**, 1168 (2010).
[27] I. Hudáky, P. Hudáky, and A. Perczel, J. Comput. Chem. **25**, 1522 (2004).
[28] I. R. Gould, W. D. Cornell, and I. H. Hillier, J. Am. Chem. Soc. **116**, 9250 (1994).
[29] A.-X. Wang and Y. Duan, J. Comput. Chem. **25**, 1699 (2004).
[30] T. J. Marrone, M. K. Gilson, and J. A. McCammon, J. Phys. Chem. **100**, 1439 (1996).
[31] Y. K. Kang, J. Phys. Chem. B **110**, 21338 (2006).
[32] Y. S. Kim, J. Wang and R. M. Hochstrasser, Phys. Chem. B **109**, 7511 (2005).
[33] J. Graf, P. H. Nguyen, G. Stock, and H. Schwalbe, J. Am. Chem. Soc. **129**, 1179 (2007).
[34] D. J. Tobias and C. L. Brooks, J. Chem. Phys. **96**, 3864 (1992).
[35] S. G. Kalko, E. Guardia, and J. A. Padró, J. Phys. Chem. B **103**, 3935 (1999).
[36] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, and M. Karplus, J. Phys. Chem. B **105**, 569 (2001).
[37] T. Head-Gordon, M. Head-Gordon, M. J. Frisch, C. L. Brooks III, and J. A. Pople, J. Am. Chem. Soc. **113**, 5989 (1991).
[38] Y. Grenie, M. Avignon, and C. Garrigoulagrang, J. Mol. Struct. **24**, 293 (1975).
[39] I. Fdez. Galván, M. L. Sánchez, M. E. Martín, F. J. Olivares del Valle, and M. A. Aguilar, Comput. Phys. Commun. **155**, 244 (2003).
[40] M. L. Sánchez, M. E. Martín, I. Fdez. Galván, F. J. Olivares del Valle, and M. A. Aguilar, J. Phys. Chem. B **106**, 4813 (2002).
[41] M. E. Martín, M. L. Sánchez, F. J. Olivares del Valle, and M. A. Aguilar, J. Chem. Phys. **113**, 6308 (2000).
[42] M. E. Martín, M. L. Sanchez, M. A. Aguilar, and F. J. Olivares del Valle, J. Mol. Struct.: THEOCHEM, **537**, 213 (2001).
[43] M. E. Martín, A. Muñoz-Losa, I. Fdez. Galván, and M. A. Aguilar, J. Chem. Phys. **121**, 3710 (2004).
[44] A. Muñoz-Losa, I. Fdez. Galván, M. E. Martín, and M. A. Aguilar, J. Phys. Chem. B **110**, 18064 (2006).
[45] I. Fdez Galván, M. A. Aguilar, and M. F. Ruiz López, J. Phys. Chem B. **109**, 23024 (2005).
[46] I. Fdez. Galván, F. J. Olivares del Valle, M. E. Martín, and M. A. Aguilar, Theor. Chem. Acc. **111**, 196 (2004).
[47] J. C. Corchado, M. L. Sánchez, and M. A. Aguilar, J. Am. Chem. Soc. **126**, 7311 (2004).
[48] C. Soriano-Correa, F. J. Olivares-del-Valle, A. Muñoz-Losa, I. Fdez. Galván, M. E. Martín, and M. A. Aguilar, J. Phys. Chem. B **114**, 8961 (2010).

[49]A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[50]M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[51]V. Luzhkov and A. Warshel, J. Comput. Chem. **13**, 199 (1992).
[52]J. Gao, J. Phys. Chem. **96**, 537 (1992).
[53]V. V. Vasilyev, A. A. Bliznyuk, and A. A. Voityuk, Int. J. Quantum Chem. **44**, 897 (1992).
[54]V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, and G. G. Ferenczy, J. Comput. Chem. **15**, 269 (1994).
[55]M. A. Thompson, E. D. Glendening, and D. Feller, J. Phys. Chem. **98**, 10465 (1994).
[56]I. Fdez. Galván, M. L. Sánchez, M. E. Martín, F. J. Olivares del Valle, and M. A. Aguilar, J. Chem. Phys. **118**, 255 (2003).
[57]T. Yamamoto, J. Phys. Chem. **129**, 244104 (2008).
[58]N. Okuyama-Yoshida, M. Nagaoka, and T. Yamabe, Int. J. Quantum Chem. **70**, 95 (1998).
[59]N. Okuyama-Yoshida, K. Kataoka, M. Nagaoka, and T. Yamabe, J. Chem. Phys. **113**, 3519 (2000).
[60]H. Hirao, Y. Nagae, and M. Nagaoka, Chem. Phys. Lett. **348**, 350 (2001).
[61]P. A. Kollman, Chem. Rev. **93**, 2395 (1993).
[62]A. E. Mark, in *Encyclopedia of Computational Chemistry*, edited by P. V. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. Schreiner (Wiley and Sons, Chichester, 1998), Vol. **2**, p. 1070.
[63]M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 98, Revision A.11.3, Gaussian, Inc., Pittsburgh, PA, 1998.
[64]K. Refson, Comput. Phys. Commun. **126**, 310 (2000).
[65]W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[66]L. E. Chirlian and M. M. Francl, J. Comput. Chem. **8**, 894 (1987).
[67]C. M. Breneman and K. B. Wiberg, J. Comput. Chem. **11**, 316 (1990).
[68]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[69]W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[70]R. Improta and V. Barone, J. Comput. Chem. **25**, 1333 (2004).
[71]See supplementary material at http://dx.doi.org/10.1063/1.3658857 for Cartesian coordinates of the alanine dipeptide, in gas phase and in water solution, for the different conformers. Some of the most representative geometric parameters in gas phase and in water solution together with the representation of the different minima can be found as well.