# Role of Solvent in Determining Conformational Preferences of Alanine Dipeptide in Water

Alexander N. Drozdov,[†] Alan Grossfield,[‡] and Rohit V. Pappu*,[†]

*Contribution from the Department of Biomedical Engineering and Center for Computational Biology, Washington University in St. Louis, Campus Box 1097, St. Louis, Missouri 63130, and Department of Biochemistry & Molecular Biophysics and Center for Computational Biology, Washington University School of Medicine, Campus Box 8231, St. Louis, Missouri 63110*

Received October 14, 2003; E-mail: pappu@biomed.wustl.edu

**Abstract:** Evidence from a variety of spectroscopic probes indicates that ($\phi$, $\psi$) values corresponding to the left-handed polyproline II helix ($P_{II}$) are preferred for short alanine-based peptides in water. On the basis of results from theoretical studies, it is believed that the observed preference is dictated by favorable peptide−solvent interactions, which are realized through formation of optimal hydrogen-bonding water bridges between peptide donor and acceptor groups. In the present study, we address this issue explicitly by analyzing the hydration structure and thermodynamics of 16 low-energy conformers of the alanine dipeptide (*N*-acetylalanine-*N'*-methylamide) in liquid water. Monte Carlo simulations in the canonical ensemble were performed under ambient conditions with all-atom OPLS parameters for the alanine dipeptide and the TIP5P model for water. We find that the number of hydrogen-bonded water molecules connecting the peptide group donor and acceptor atoms has no effect on the solvation thermodynamics. Instead, the latter are determined by the work done to fully hydrate the peptide. This work is minimal for conformations that are characterized by a minimal overlap of the primary hydration shells around the peptide donor and acceptor atoms. As a result, peptide−solvent interactions favor "compact" conformations that do not include $P_{II}$-like geometries. Our main conclusion is that the experimentally observed preference for $P_{II}$ does not arise due to favorable direct interactions between the peptide and water molecules. Instead, the latter act to unmask underlying conformational preferences that are a consequence of minimizing intrapeptide steric conflicts.

## Introduction

Recent advances in optical spectroscopy allow close scrutiny of the ensemble average of conformations accessible to peptides in aqueous environments.[1−9] Of specific interest are short alanine-based peptides because they are good model systems to study local conformational preferences of unfolded polypeptides under folding conditions.[1,6,9−12] Evidence based on a

variety of spectroscopic probes indicates that the dominant conformer for such peptides takes on backbone dihedral angles corresponding to the left-handed polyproline II helix ($P_{II}$).[1,3−12] This and other observations have helped rekindle interest[12−15] in an idea originally put forth by Tiffany and Krimm.[16,17] According to their hypothesis, unfolded proteins are a concatenation of segments that fluctuate between two types of conformational states, namely, short stretches of the $P_{II}$ helix motif and a "wide sampling of a standard dipeptide energy map".[17]

The present work is focused on one of the simplest peptides, the alanine dipeptide, shown in Figure 1. Conformational preferences of the alanine dipeptide in water have been studied by a variety of experimental[3,9] and theoretical techniques.[18−31] On the basis of results of these studies, it is now commonly

[†] Department of Biomedical Engineering, Washington University in St. Louis.

[‡] Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine.

(1) Eker, F.; Cao, X. L.; Nafie, L.; Huang, Q.; Schweitzer-Stenner, R. *J. Phys. Chem. B* **2003**, *107*, 358−365.
(2) Ohnishi, S.; Shortle, D. *Proteins: Struct. Funct. Genet.* **2003**, *50*, 546−551.
(3) Weise, C. F.; Weisshaar, J. C. *J. Phys. Chem. B* **2003**, *107*, 3265−3277.
(4) Barron, L. D.; Blanch, E. W.; Hecht, L. *Adv. Protein Chem.* **2002**, *62*, 51−90.
(5) Keiderling, T. A.; Xu, Q. *Adv. Protein Chem.* **2002**, *62*, 111−161.
(6) Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K. *J. Am. Chem. Soc.* **2001**, *123*, 9628−9633.
(7) Gnanakaran, S.; Hochstrasser, R. M. *J. Am. Chem. Soc.* **2001**, *123*, 12886−12898.
(8) Woutersen, S.; Hamm, P. *J. Phys. Chem. B* **2000**, *104*, 11316−11320.
(9) Poon, C.-D.; Samulski, E. T.; Weise, C. F.; Weisshaar, J. C. *J. Am. Chem. Soc.* **2000**, *122*, 5642−5643.
(10) Woody, R. W. *Adv. Biophys. Chem.* **1992**, *2*, 37−79.
(11) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190−9195.
(12) Shi, Z.; Woody, R. W.; Kallenbach, N. R. *Adv. Protein Chem.* **2002**, *62*, 163−240.

(13) Rucker, A. L.; Creamer, T. P. *Protein Sci.* **2002**, *11*, 980−985.
(14) Ferreon, J. C.; Hilser, V. J. *Protein Sci.* **2003**, *12*, 447−457.
(15) Ding, L.; Chen, K.; Santini, P. A.; Shi, Z. S.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2003**, *125*, 8092−8093.
(16) Tiffany, M. L.; Krimm, S. *Biopolymers* **1969**, *8*, 347−359.
(17) Krimm, S.; Tiffany, M. L. *Israel J. Chem.* **1974**, *12*, 189−200.
(18) Pettitt, B. M.; Karplus, M. *J. Phys. Chem.* **1988**, *92*, 3994−3997.
(19) Anderson, A. G.; Hermans, J. *Proteins: Struct. Func. Genet.* **1988**, *3*, 262−265.
(20) Tobias, D. J.; Brooks, C. L. *J. Phys. Chem.* **1992**, *96*, 3864−3870.
(21) Yang, J. H.; Shin, J. K.; Jhon, M. S. *J. Mol. Struct.* **1992**, *268*, 169−179.
(22) Brooks, C. L.; Case, D. A. *Chem. Rev.* **1993**, *93*, 2487−2502.
(23) Bartels, C.; Karplus, M. *J. Comput. Chem.* **1997**, *18*, 1450−1462.
(24) Resat, H.; Maye, P. V.; Mezei, M. *Biopolymers* **1997**, *41*, 73−81.
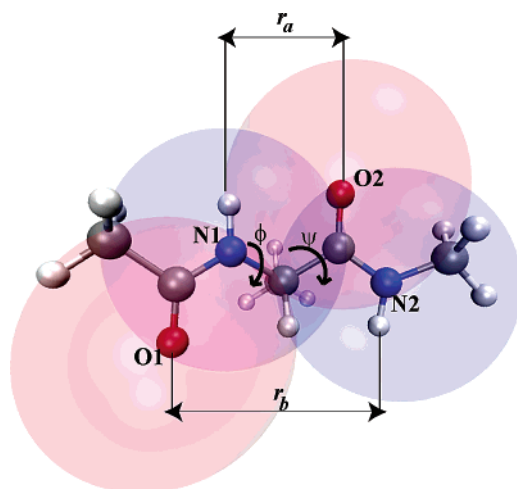
**Figure 1.** Ball and stick model for the alanine dipeptide. Labels O1, O2, N1, and N2 are for the dipeptide group acceptor and donor atoms. Red and blue spheres around the oxygen and nitrogen atoms, respectively, are used to delineate the primary hydration shells around these atoms. Distances $r_a$ and $r_b$ specify the extent of overlap between the associated shells.

recognized that the preference of aqueous alanine for $P_{II}$-like conformers is determined by favorable peptide−solvent interactions.[9,10,12] The validity of this assertion follows from the fact that $P_{II}$, which is the global minimum on the hydrated free energy surface, is not even a local minimum on the gas-phase potential energy surface.[20,22,28] The question of interest is, How does water influence peptide geometry? Han et al.[31] studied this issue using density functional theory calculations. A single alanine dipeptide and four coordinated water molecules were placed in a spherical Onsager cavity to mimic the effect of peptide solvation. Two principal results of this study are as follows: First, the four water molecules form single- and double-water bridges between the peptide donor and acceptor groups, and, second, the $P_{II}$ conformer is the most stable one. On the basis of these results, Poon et al.[9] surmised that formation of optimal hydrogen-bonding water bridges is likely to be a key determinant of the structure of alanine dipeptide in aqueous solution. A similar view has been reiterated in the recent review of Shi et al.[12] However, this conjecture contradicts results of Pappu and Rose,[32] who showed that the preference of alanine dipeptide for $P_{II}$ and other $\beta$-like conformers can be explained solely by minimization of local steric conflicts.

In the present paper, the issue of what interactions are responsible for the predominance of $P_{II}$ is studied explicitly without relying on simplified models.

## 2. Methods

The results presented in this work were obtained from molecular dynamics and Monte Carlo simulations.[33−39] The molecular dynamics

(25) Smart, J. L.; Marrone, T. J.; McCammon, J. A. *J. Comput. Chem.* **1997**, *18*, 1750−1759.
(26) Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. B* **1998**, *102*, 3637−3641.
(27) Tazaki, K.; Shimizu, K. *J. Phys. Chem. B* **1998**, *102*, 6419−6424.
(28) Smith, P. E. *J. Chem. Phys.* **1999**, *111*, 5568−5579.
(29) Apostolakis, J.; Ferrara, P.; Caflisch, A. *J. Chem. Phys.* **1999**, *110*, 2099−2108.
(30) Hu, H.; Elstner, M.; Hermans, J. *Proteins: Struct. Funct. Genet.* **2003**, *50*, 451−463.
(31) Han, W.-G.; Jalkanen, K. J.; Elstner, M.; Suhai, S. *J. Phys. Chem. B* **1998**, *102*, 2587−2602.
(32) Pappu, R. V.; Rose, G. D. *Protein Sci.* **2002**, *11*, 2437−2455.
(33) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(34) Owicki, J. C.; Scheraga, H. A. *Chem. Phys. Lett.* **1977**, *47*, 600−602.

simulations were carried out for a flexible alanine dipeptide (Figure 1) in water in order to calculate the probability distribution $P(\phi, \psi)$ for the dihedral angles. This distribution was used to identify a set of low-energy conformations that characterize important regions of the Ramachandran map.[40] The distribution of water molecules around alanine dipeptide in these selected rigid conformations was then studied by the Metropolis Monte Carlo method.[33]

Both sets of simulations were performed in a canonical ensemble at ambient conditions, $T = 298$ K and $\rho = 1$ g cm$^{-3}$. A single dipeptide and 256 water molecules were placed in a cubic box of side 20 Å with periodic boundary conditions imposed. Parameters from the all-atom OPLS force field[41] were used for the alanine dipeptide, and the five-site TIP5P water model[42] was used for the solvent. All long-range interactions were handled by use of spherical cutoffs set at half the box length. The nonbonded terms were smoothly varied from their standard value at 9.0 Å to zero at 10 Å via a tapering scheme based on a polynomial switching function.[37,39]
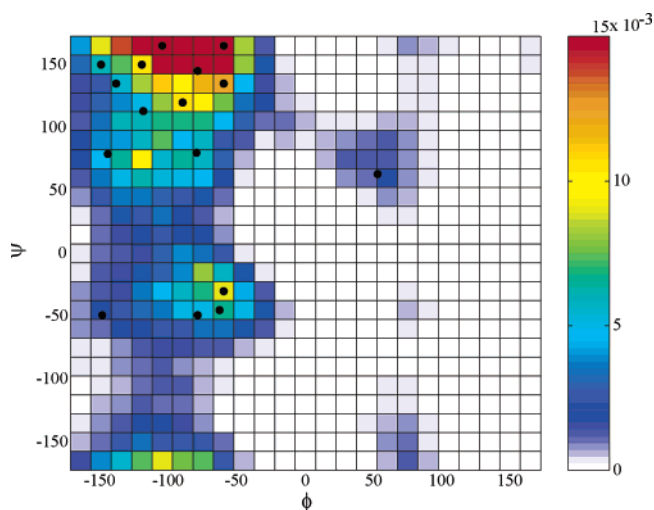
**2.1. Molecular Dynamics Simulations.** The dynamics were described by Langevin equations of motion to simulate a canonical ensemble.[35] The velocity Verlet algorithm was employed to integrate these equations of motion with a time step of 1 fs.[36] The internal geometry of water molecules was held rigid by the RATTLE procedure.[38]

To construct the probability distribution $P(\phi, \psi)$, 288 independent trajectories were generated. For a given trajectory, a harmonic potential of the form $U_r(\phi, \psi; \phi_{om}, \psi_{on}) = (k/2)[(\phi - \phi_{om})^2 + (\psi - \psi_{on})^2]$ was applied to restrain the backbone angles of the dipeptide to a particular region centered at $(\phi_{om}, \psi_{on})$. The force constant $k$ was 0.02 kcal/(mol deg$^2$), while $\phi_{om}$ and $\psi_{on}$ were generated on two 30° × 30° grids. The two grids were shifted with respect to each other by 15° such that one grid was given by $(\phi_{om}, \psi_{on}) = (-180° + 30°m, -180° + 30°n)$ and the other by $(\phi_{om}, \psi_{on}) = (-165° + 30°m, -165° + 30°n)$ with $0 \leq m, n \leq 11$. Each trajectory was run for 100 ps, and the first 40 ps was discarded. The starting configuration for each simulation was first preequilibrated and then refined by energy minimization in the presence of the restraining potential. A total of $2 \times 12^2 = 288$ biased probability distributions, $P_i(\phi, \psi)$, were generated. The unbiased probability distribution $P(\phi, \psi)$ for the alanine dipeptide in water was then constructed by the weighted histogram analysis method.[43−45]

To quantify the effect of long-range interactions, we calculated the probability distribution $P(\phi, \psi)$ using two different methods for evaluating Coulomb interactions: spherical cutoffs and Ewald sums.[37,38] The distribution functions so obtained were found to agree with each other within statistical error. This is easy to understand: both the alanine dipeptide and water molecules are small electrically neutral entities. Separations between charges within these molecules are relatively short. At distances larger than these separations, electrostatic interactions reduce to dipole−dipole interactions, which fall off much faster than

(35) Allen, M. P. *Mol. Phys.* **1980**, *40*, 1073−1087.
(36) Guarnieri, F.; Still, W. C. *J. Comput. Chem.* **1994**, *15*, 1302−1310.
(37) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Oxford Science Publications: Oxford, 1987.
(38) Frenkel, D.; Smit, B. *Understanding molecular simulation: From algorithms to applications*; Academic Press: New York, 2002.
(39) We used the TINKER modeling package, Version 3.9, (http:// dasher.wustl.edu/∼tinker) for force field parameters and routines for energy evaluation, molecular dynamics simulations and weighted histogram analysis. For the Monte Carlo calculations, we developed custom routines based on the standard Metropolis algorithm.[33]
(40) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95−99.
(41) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *117*, 11225−11236. Maxwell, D. S.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1995**, *16*, 984−1010.
(42) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910−8922.
(43) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(44) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339−1350.
(45) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275−282.

**Figure 2.** Two-dimensional probability density plot obtained for aqueous alanine dipeptide from the molecular dynamics simulations. The adjacent color bar is used to identify regions of low versus high population. Locations of the conformers, listed in Table 1, are indicated by solid circles.

**Table 1.** List of Alanine Dipeptide Conformers Used in Monte Carlo Simulations

| conformer name | $(\phi, \psi)$ values (deg) | conformer name | $(\phi, \psi)$ values (deg) |
|---|---|---|---|
| $C_5$ | $(-150, 150)$ | $P_4$ | $(-60, 165)$ |
| $\beta_A$ | $(-139, 135)$ | $P_5$ | $(-60, 135)$ |
| $\beta_P$ | $(-119, 113)$ | $C_{7eq}$ | $(-80, 80)$ |
| $\beta'$ | $(-90, 120)$ | $\alpha_P$ | $(-149, -49)$ |
| $\beta''$ | $(-120, 150)$ | $\alpha_R$ | $(-79, -49)$ |
| $sP_{II}$ | $(-145, 79)$ | $\alpha'_R$ | $(-63, -45)$ |
| $P_{II}$ | $(-79, 145)$ | $\alpha''_R$ | $(-60, -30)$ |
| $P_3$ | $(-105, 150)$ | $\alpha_L$ | $(53, 63)$ |

charge−charge interactions. On the basis of our results, we conclude that for the particular problem of the alanine dipeptide in water, Ewald sums can be replaced with spherical cutoffs without introducing significant errors.

**2.2. Monte Carlo Simulations.** We performed simulations of aqueous alanine dipeptide in 16 different conformations. This helped us understand how hydration influences conformational preferences of alanine dipeptide. Our main interest is in the distribution of water molecules in the immediate vicinity of the peptide. Consequently, a "smart" Monte Carlo method, also known as preferential sampling,[34,37] was employed. For each alanine dipeptide conformer, 50 uncorrelated trajectories were generated. Each trajectory started from a distinct equilibrium configuration for the peptide−solvent system and comprised $7 \times 10^6$ trial moves. A trial move involved a random rotation and translation of a randomly chosen water molecule. Translation refers to a displacement of the center of mass of the chosen water molecule, and rotation corresponds to perturbations of the three independent Euler angles based on the so-called *xyz* convention.[46] The amplitudes of these random rotations and translations were fixed such that the acceptance ratio was approximately 40%. Every $10^3$ accepted moves, a peptide−solvent configuration was saved. Thus, a total of $1.2 \times 10^5$ independent peptide−solvent configurations were generated for each alanine dipeptide conformation. Thermodynamic averages were obtained by averaging over all saved configurations, and block averages over separate runs were used to estimate statistical errors.

## 3. Results and Discussion

The main goal of our study is to understand the origin of conformational preferences for alanine dipeptide in water as computed via the molecular dynamics simulations described in section 2.1. Ideally, such a study should be carried out by calculating the distribution of solvent molecules around dipeptide conformers generated on a uniform $(\phi, \psi)$ grid. However, a large fraction of $(\phi, \psi)$ space is inaccessible due to steric conflicts.[40] An alternative that we use here is to restrict the proposed analysis to conformations from low-energy regions on the hydrated free energy surface.

**3.1. Identification of Important Conformations.** In Figure 2 we present the equilibrium distribution of the dihedral angles,

$P(\phi, \psi)$, of alanine dipeptide in water. There are two dominant basins on this map: the $\beta$- and $\alpha$-regions located in the left top and bottom quadrants, respectively. The maximum of the distribution $P(\phi, \psi)$ corresponds to a $P_{II}$-like conformation with the window centered at $\phi = -82.5°$ and $\psi = 152.5°$. Since $P(\phi, \psi)$ was generated on a coarse grid, it is difficult to identify the precise locations of minima on the hydrated free energy surface. We overcome this difficulty by choosing a set of representative conformations from highly populated regions of the Ramachandran map. The selected conformations are shown in Figure 2 and listed in Table 1. Our choice reflects the preference of alanine dipeptide for sterically allowed conformations that expose all four functional groups (O1, N1, O2, and N2 in Figure 1) to water.

Furthermore, to simplify our presentation, we identify a subset of five representative conformations to be used in the subsequent illustrations. This subset includes the canonical $P_{II}$ conformation, which is in the neighborhood of the global minimum on the hydrated free energy surface; $C_{7eq}$, the preferred conformation in gas phase;[20] $\alpha_L$ and $\alpha'_R$, dominant conformations from the top right and bottom left quadrants of the Ramachandran map; and $\beta_A$, a conformation corresponding to $(\phi, \psi)$ values of canonical $\beta$-strands. These conformations were chosen to adequately represent variations in the solvation properties revealed by studying the set of 16 conformations. Results for the latter are summarized in Table 2.

**3.2. Hydration Structure.** Conformation-dependent differences in the hydration of alanine dipeptide are mainly controlled by four peptide group donor and acceptor atoms that can form hydrogen bonds to water. Hydration of other (nonpolar) groups was found to be rather insensitive to the peptide geometry and hence will not be discussed here. First we study the hydration of alanine dipeptide in the 16 conformations, using the standard approach based on the atomic radial distribution function.[47,48] The latter, a basic quantity in theory of liquids, provides structural information.[49] In particular, the position of the first peak in a radial distribution function specifies the nearest-neighbor distance, the peak width indicates the fluctuations from this value, and the area under the peak provides an estimate of the number of nearest neighbors. We calculated the radial distribution functions for water oxygen atoms around the peptide group donor [$g_{N1}(r)$ and $g_{N2}(r)$] and acceptor [$g_{O1}(r)$ and $g_{O2}(r)$] atoms. Results for the five representative conformers are shown in Figure 3. The radial distribution functions are seen to be relatively insensitive to the peptide geometry in the sense that the different curves are roughly similar to each other within

(46) Goldstein, H. *Classical Mechanics*, 2nd ed.; Addison-Wesley: Reading, MA, 1980.

(47) Rossky, P. J.; Karplus, M. *J. Am. Chem. Soc.* **1979**, *101*, 1913−1937.
(48) Cui, Q.; Smith, V. H. *J. Chem. Phys.* **2003**, *118*, 279−290.
(49) Hansen, J.-P.; McDonald, I. R. *Theory of simple liquids*; Academic Press: New York, 1986.

***Table 2.*** Summary of Results from Monte Carlo Simulations[a]

| conformer name | $\langle n_a \rangle$[b] | $\langle n_b \rangle$[c] | $\langle n_{min} \rangle$[d] | $\sigma_i$[e] | $\langle \Delta U_{ps}{}^i \rangle$[f] | $\Delta U_P{}^i$[g] | $\Delta U_i$[h] | SASA[i] |
|---|---|---|---|---|---|---|---|---|
| $C_5$ | $1.4 \pm 0.1$ | $3.4 \pm 0.1$ | $4.8 \pm 0.1$ | $4.5 \pm 0.1$ | $5.1 \pm 0.5$ | $-1.0$ | $4.1 \pm 0.5$ | 359 |
| $\beta_A$ | $1.8 \pm 0.1$ | $3.1 \pm 0.2$ | $4.9 \pm 0.3$ | $4.4 \pm 0.2$ | $4.6 \pm 0.6$ | $-0.5$ | $4.1 \pm 0.6$ | 360 |
| $\beta_P$ | $2.8 \pm 0.2$ | $2.4 \pm 0.2$ | $5.2 \pm 0.3$ | $4.3 \pm 0.2$ | $4.6 \pm 0.6$ | $-0.1$ | $4.5 \pm 0.6$ | 361 |
| $\beta'$ | $2.5 \pm 0.2$ | $3.1 \pm 0.1$ | $5.6 \pm 0.2$ | $4.3 \pm 0.2$ | $3.6 \pm 0.5$ | $-0.7$ | $2.9 \pm 0.3$ | 361 |
| $\beta''$ | $3.4 \pm 0.2$ | $2.6 \pm 0.1$ | $6.0 \pm 0.2$ | $4.3 \pm 0.1$ | $3.2 \pm 0.3$ | $-0.6$ | $2.6 \pm 0.5$ | 359 |
| $sP_{II}$ | $3.4 \pm 0.2$ | $2.8 \pm 0.1$ | $6.2 \pm 0.3$ | $4.5 \pm 0.2$ | $2.8 \pm 0.6$ | $1.2$ | $4.0 \pm 0.6$ | 359 |
| $P_{II}$ | $3.6 \pm 0.2$ | $3.3 \pm 0.2$ | $6.9 \pm 0.3$ | $4.6 \pm 0.2$ | $0.0 \pm 0.6$ | $0.0$ | $0.0 \pm 0.6$ | 358 |
| $P_3$ | $3.1 \pm 0.1$ | $3.5 \pm 0.1$ | $6.6 \pm 0.1$ | $4.5 \pm 0.2$ | $0.5 \pm 0.4$ | $0.1$ | $0.6 \pm 0.4$ | 359 |
| $P_4$ | $4.0 \pm 0.3$ | $3.9 \pm 0.3$ | $7.9 \pm 0.5$ | $4.9 \pm 0.2$ | $-4.5 \pm 0.5$ | $4.9$ | $0.4 \pm 0.5$ | 355 |
| $P_5$ | $3.8 \pm 0.4$ | $3.4 \pm 0.2$ | $7.2 \pm 0.5$ | $4.7 \pm 0.2$ | $-0.7 \pm 0.5$ | $2.1$ | $1.6 \pm 0.5$ | 355 |
| $C_{7eq}$ | $3.7 \pm 0.1$ | $0.0$ | $3.7 \pm 0.1$ | $4.2 \pm 0.2$ | $10.2 \pm 0.4$ | $-2.2$ | $8.0 \pm 0.4$ | 354 |
| $\alpha_P$ | $5.3 \pm 0.2$ | $5.4 \pm 0.2$ | $10.7 \pm 0.3$ | $4.6 \pm 0.1$ | $0.0 \pm 0.6$ | $4.2$ | $4.2 \pm 0.6$ | 358 |
| $\alpha_R$ | $5.9 \pm 0.2$ | $4.2 \pm 0.5$ | $10.1 \pm 0.4$ | $4.8 \pm 0.2$ | $-2.4 \pm 0.3$ | $3.6$ | $1.2 \pm 0.3$ | 357 |
| $\alpha'_R$ | $5.8 \pm 0.2$ | $4.5 \pm 0.3$ | $10.3 \pm 0.4$ | $4.9 \pm 0.3$ | $-3.3 \pm 0.5$ | $4.2$ | $0.9 \pm 0.5$ | 354 |
| $\alpha''_R$ | $5.8 \pm 0.2$ | $3.9 \pm 0.4$ | $9.7 \pm 0.4$ | $4.8 \pm 0.2$ | $-1.4 \pm 0.4$ | $5.2$ | $3.8 \pm 0.4$ | 352 |
| $\alpha_L$ | $6.0 \pm 0.2$ | $5.1 \pm 0.2$ | $11.1 \pm 0.2$ | $5.1 \pm 0.2$ | $-5.9 \pm 0.6$ | $11.0$ | $5.1 \pm 0.6$ | 347 |

[a] All errors were evaluated by using block averages as described in section 2.2. [b] $\langle n_a \rangle$, average minimal path between sites N1 and O2. [c] $\langle n_b \rangle$, average minimal path between sites O1 and N2. [d] $\langle n_{min} \rangle = \langle n_a \rangle + \langle n_b \rangle$. [e] $\sigma_i$, width of the distribution $P(U_{PS}{}^i)$ of peptide−solvent potential energies in kilocalories per mole. [f] $\langle \Delta U_{PS}{}^i \rangle = \langle U_{PS}{}^i \rangle - \langle U_{PS}{}^{P_{II}} \rangle$, ensemble average of peptide−solvent potential energies, in kilocalories per mole, relative to that for $P_{II}$. [g] $\Delta U_P{}^i = U_P{}^i - U_P{}^{P_{II}}$, gas-phase potential energy in kilocalories per mole relative to that for $P_{II}$. [h] $\Delta U_i = \langle \Delta U_{PS}{}^i \rangle + \Delta U_P{}^i$ in kilocalories per mole. [i] SASA, solvent-accessible surface area in square angstroms.
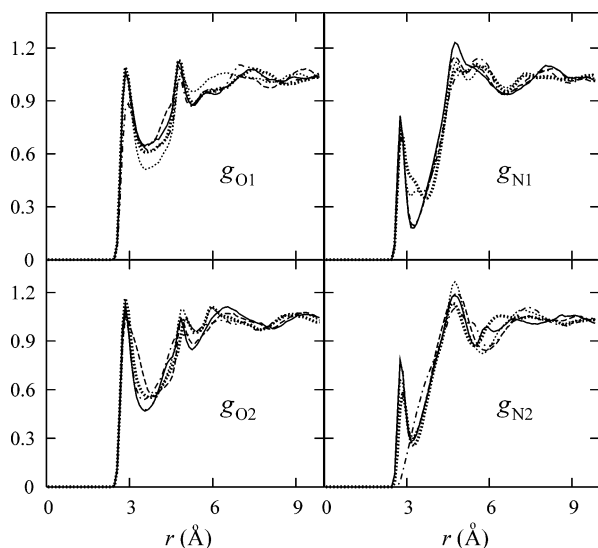


***Figure 3.*** Radial distribution functions of water molecules around the dipeptide group donor (N1, N2) and acceptor (O1, O2) atoms. The different line styles represent the results for five different conformations: $P_{II}$ (solid lines), $\beta_A$ (dashed lines), $C_{7eq}$ (dotted−dashed lines), $\alpha_L$ (dotted lines), and $\alpha'_R$ (thick dashed lines).
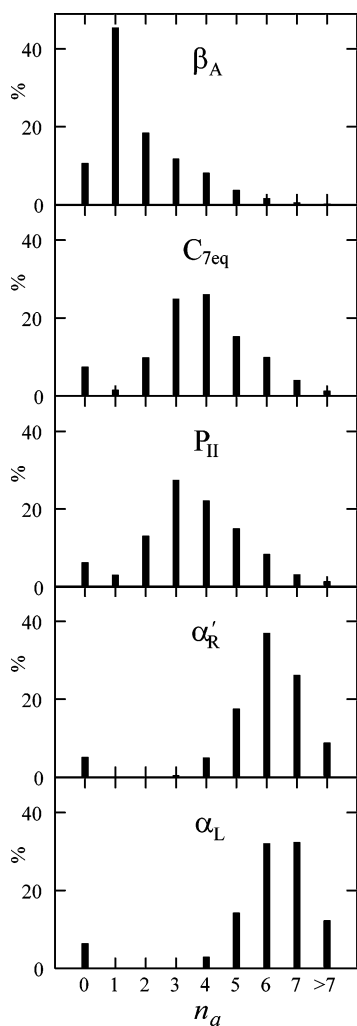
statistical errors. The reason for their similarity is peptide−solvent hydrogen bonding, due to which there can be on average one and two nearest-neighbor water molecules around the peptide group donor and acceptor atoms, respectively. This is true for all conformations except $C_{7eq}$. In the latter there is an intramolecular hydrogen bond between sites O1 and N2; consequently, these sites are not fully hydrated. Beyond computation of primary hydration shell occupancies, no other useful discriminatory information regarding the conformational dependence of the solvation structure can be obtained on the basis of radial distribution functions alone. In principle, one could calculate three-dimensional spatial and orientational distribution functions.[50] However, in an attempt to make contact with the density functional theory calculations of Han et al.[31] we pursue an alternative approach to quantify hydration structures.

The peculiar physical and chemical properties of liquid water are related to its open, tetrahedral structure that is realized through a dynamic network of hydrogen bonds.[51−54] In the vicinity of the peptide, the bonding properties of water are likely to be influenced by the peptide geometry. To quantify this dependence, we define a path as the number of hydrogen-bonded water molecules required to connect pairs of peptide group donor and acceptor sites, i.e., either N1 and O2 or N2 and O1.[55] Our interest is in the minimal path, $n_{min}$, for a given peptide−solvent configuration because it characterizes the local water network around the peptide. This quantity is a sum of two values $n_a$ and $n_b$ which are the minimal numbers of hydrogen-bonded water molecules connecting the adjacent (backbone) peptide sites N1 and O2 ($n_a$), and the blocking group donor and acceptor atoms N2 and O1 ($n_b$). The instantaneous value of $n_{min}$ varies from one snapshot to another. Accordingly, we introduce a new measure, the average minimal path $\langle n_{min} \rangle$, which is obtained by averaging over all peptide−solvent configurations saved for a given peptide conformation.

To calculate the minimal path we used a standard distance-based criterion to identify hydrogen bonds.[53,56−58] According to this criterion, two water molecules or a water molecule and carbonyl oxygen (O1 or O2) are considered to be hydrogen-bonded if the oxygen−oxygen distance is no greater than 3.3 Å. Similarly, the amide group is considered to be in a hydrogen bond if the distance between a water oxygen and amide nitrogen (N1 or N2) is no greater than 3.3 Å. The chosen distance of 3.3 Å corresponds to the location of the first minimum in the associated radial distribution functions (see Figure 3).
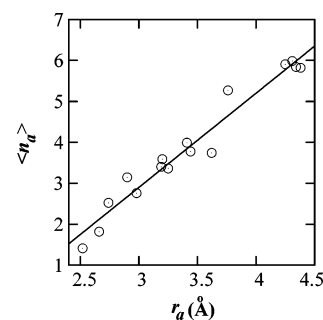
(50) Bergman, D. L. *Chem. Phys.* **2000**, *253*, 267−282.

(51) Geiger, A.; Stillinger, F. H.; Rahman, A. *J. Chem. Phys.* **1979**, *70*, 4185−4193.
(52) Stanley, H. E.; Blumberg, R. L. *Phys. Rev. B* **1983**, *28*, 1626−1629.
(53) Robinson, G. W.; Zhu, S.-B.; Singh, S.; Evans, M. W. *Water in Biology, Chemistry and Physics*; World Scientific: Singapore, 1994.
(54) Teixeira, J.; Luzar, A. In *Hydration Processes in Biology*; Bellisent-Funel, M.-C., Ed; IOS press: London, 1999.
(55) Sreerama, N.; Woody, R. W. *Proteins: Struct. Func. Genet.* **1999**, *36*, 400−406.
(56) Mezei, M.; Beveridge, D. L. *J. Chem. Phys.* **1981**, *74*, 622−632.
(57) Xu, H.; Berne, B. J. *J. Phys. Chem. B* **2001**, *105*, 11929−11932.
(58) Kosztolányi, T.; Bakó, I.; Pálinkas, G. *J. Chem. Phys.* **2003**, *118*, 4546−4555.

**Figure 4.** Percentage distribution of minimal paths $n_a$ for different alanine dipeptide conformers indicated by the legends. $n_a = 0$ corresponds to the null path.



**Figure 5.** Minimal path $\langle n_a \rangle$ as a function of the distance $r_a$. Circles denote results from Monte Carlo simulations; solid line shows the results of linear regression.

The average minimal path defined above is given by the sum of the two numbers, $\langle n_{min} \rangle = \langle n_a \rangle + \langle n_b \rangle$. Results obtained for $\langle n_a \rangle$, $\langle n_b \rangle$, and $\langle n_{min} \rangle$ are presented in Table 2. To facilitate interpretation of these results, the distribution of $n_a$ for the five representative conformers is shown in Figure 4. The finite box-size limits us to considering paths that have no more than seven water molecules. A null path, $n_a = 0$, occurs when either N1 or O2 have no water molecules in their primary hydration shells. Similarly, $n_b = 0$ implies that there is no water molecule in the primary hydration shell around either N2 or O1. Unlike the radial distribution functions shown in Figure 3, the distribution of minimal paths exhibits strong conformational dependence. The mean value $\langle n_a \rangle$ varies between 1.4 for $C_5$ and 6 for $\alpha_L$. These variations in $\langle n_a \rangle$ are associated with the variations in the distance between the backbone carbonyl oxygen and amide hydrogen atoms ($r_a$ in Figure 1). The latter follows from Figure 5, which depicts the mean value $\langle n_a \rangle$ as a function of the distance $r_a$. Linear regression analysis indicates that we ought to expect 2.3 water molecules in the minimal path between sites N1 and O2 for every angstrom of separation between adjacent amide hydrogen and carbonyl oxygen atoms. Similar results were obtained for $\langle n_b \rangle$ and $r_b$. The only exception is $C_{7eq}$. In this conformation, the blocking group donor (N2) and acceptor (O1)

atoms are in a hydrogen bond with each other and hence they cannot be connected by a path with $n_b \neq 0$.

**3.3. Solvation Thermodynamics.** In this section, we address the issue of peptide–solvent interactions. Specifically, do conformation-dependent minimal paths between peptide sites influence solvation thermodynamics, and to what extent is the preference of aqueous alanine dipeptide for $P_{II}$ dictated by favorable peptide–solvent interactions? To answer these questions, one needs to analyze the origin of the Helmholtz free-energy difference, $\Delta A$, between the $P_{II}$ basin and other regions of the Ramachandran map (Figure 2). Since our simulations were performed for rigid conformers, we base our detailed (quantitative) analysis on the potential-energy contribution to the free energy, $\Delta U_i = U_i - U_{P_{II}}$, whereas the associated change of the solvent entropy around the peptide will be discussed only qualitatively. Furthermore, to be able to distinguish effects of intrapeptide and peptide–solvent interactions, we split the potential-energy difference into two terms, $\Delta U_i = \Delta U_P{}^i + \langle \Delta U_{PS}{}^i \rangle$, where $\Delta U_P{}^i = U_P{}^i - U_P{}^{P_{II}}$ and $\langle \Delta U_{PS}{}^i \rangle = \langle U_{PS}{}^i \rangle - \langle U_{PS}{}^{P_{II}} \rangle$ are the relative changes in the gas-phase and average peptide–solvent potential energies, respectively. The former were calculated from analytical expressions, while the latter were evaluated by Monte Carlo simulations.

Results of our calculations are shown in Table 2 along with those for the solvent-accessible surface area. A striking finding is that $P_{II}$, which has the lowest free energy of the conformers with significant populations in solution (see top left quadrant of the Ramachandran plot in Figure 2), is *not* the global minimum with respect to solvation. Instead, the most solvated conformation is $\alpha_L$ and the least solvated is $C_{7eq}$, while $P_{II}$ lies between these two extremes. The data shown in Table 2 also establish that within the top-left quadrant, the dominant region of $(\phi, \psi)$ space, favorable peptide–solvent interactions prefer $P_{II}$-like geometries over $\beta$-like conformations. The information contained in Table 2 is further analyzed in Figure 6, which shows the probability distribution of peptide–solvent potential energies $P(U_{PS}{}^i)$ for the five representative conformers. This distribution is Gaussian in shape and can therefore be defined by the first two cumulants, the mean $\langle \Delta U_{PS}{}^i \rangle$ and variance $\langle (\Delta U_{PS}{}^i - \langle \Delta U_{PS}{}^i \rangle)^2 \rangle$. The mean (also known as the solvation energy) characterizes the strength of the peptide–solvent interactions. As evidenced by Figure 6 and Table 2, low-energy peptide–solvent interactions are realized for "compact" conformers ($\alpha_L$, $P_4$, and $\alpha'_R$), which are defined by greater distances between the peptide donor and acceptor sites and therefore allow water more "room" in the primary hydration shells around these sites than does $P_{II}$. The width of the distribution, $\sigma_i =$
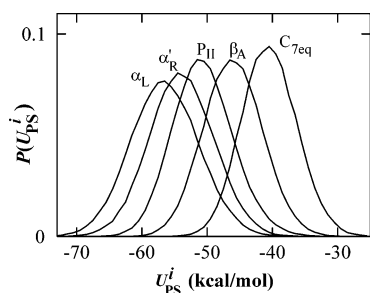
**Figure 6.** Probability distribution of peptide−solvent potential energies for five representative dipeptide conformers indicated by the legends.
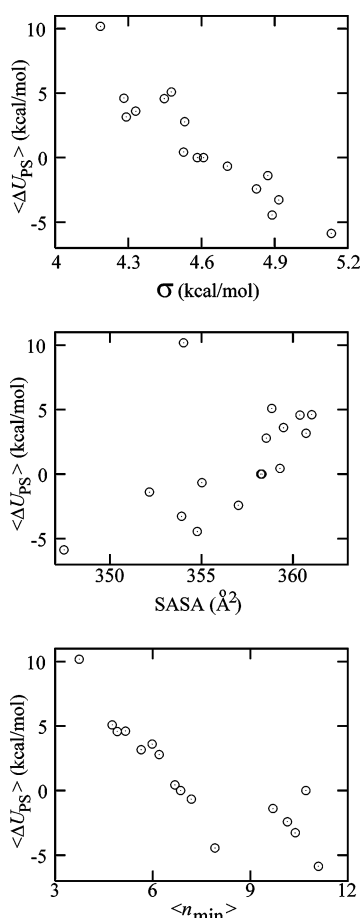


**Figure 7.** Average peptide−solvent potential energy $\langle \Delta U_{PS}^i \rangle$ versus (a, top) the magnitude of energy fluctuations $\sigma_i$, (b, middle) solvent-accessible surface area, and (c, bottom) minimal path $\langle n_{min} \rangle$.

$\sqrt{\langle (\Delta U_{PS}^i - \langle \Delta U_{PS}^i \rangle)^2 \rangle}$, determines the magnitude of fluctuations around $\langle \Delta U_{PS}^i \rangle$ for a given conformer, and this in turn is a measure of the entropic contribution to the solvation thermodynamics. The greater the width of fluctuations, the greater the contribution from solvent entropy to the free energy. The data presented in Table 2 and Figure 7a show that $\sigma_i$ is anticorrelated with $\langle \Delta U_{PS}^i \rangle$. In other words, conformers that are favorable energetically are also favorable entropically from the standpoint of peptide−solvent interactions. Indeed, a maximal width of 5.1 kcal/mol is attained for the most solvated conformer $\alpha_L$, and the least solvated conformer $C_{7eq}$ is characterized by a minimal width of 4.2 kcal/mol. However, the conformational dependence of the fluctuation width, while significant, shows a much smaller change than does the average peptide−solvent

energy. On the basis of the above observations, we conclude that the preference for $P_{II}$ *cannot* be attributed to favorable peptide−solvent interactions.

We continue our analysis by comparing the average peptide−solvent potential energy to the solvent-accessible surface area.[59] The latter is an empirical measure often used to quantify the degree to which a peptide favors interactions with the solvent over interactions with itself.[60] Specifically, it is believed that the greater the solvent-accessible surface area, the greater the affinity of the peptide for the solvent. We calculated the solvent-accessible surface area of the alanine dipeptide in the 16 conformations using a probe radius of 1.4 Å and the force field van der Waals radii for the peptide atoms. As evidenced by Table 2 and Figure 7b, the conformational dependence of the solvent-accessible surface area is contradictory to that of the average peptide−solvent potential energy. A lowest value of 347 Å$^2$ is obtained for the most solvated conformer $\alpha_L$, and one of the least solvated conformers, $\beta_P$, is characterized by the largest solvent-accessible surface area of 361 Å$^2$.

To clarify why compact conformations characterized by smaller values of the solvent-accessible surface area are more favorably solvated, we need to identify an appropriate characteristic of peptide geometry that is a major determinant of the magnitude of peptide−solvent interactions. On the basis of the data of Table 2 and Figure 7c, we notice an obvious correlation between the average minimal path $\langle n_{min} \rangle$ and the average peptide−solvent potential energy $\langle \Delta U_{PS}^i \rangle$. Indeed, the lowest energy is obtained for $\alpha_L$, which is characterized by the largest minimal path of 11.1. Conversely, $C_{7eq}$, which is the obvious outlier from the standpoint of peptide−solvent interactions, is characterized by the smallest minimal path of 3.7. The average minimal path associated with $P_{II}$, 6.9, is between the two extremes. These observations suggest that the strength of peptide−solvent interactions is tied to the average minimal path, i.e., the larger the path, the lower the energy of peptide−solvent interactions. To test the validity of this hypothesis, we calculated, for each conformer $i$, a set of probability distribution functions of peptide−solvent potential energies, $P(U_{PS}^i | n_a)$. Each of these functions is associated with a specific value of the minimal path, $n_a = 0, ..., 7$. The results obtained are shown in Figure 8. Similar results were obtained in terms of the minimal path $n_b$ that connects the blocking group sites O1 and N2. The most striking observation is that, for a given conformer, the distribution functions $P(U_{PS}^i | n_a > 0)$ associated with nonzero values of $n_a$ coincide with each other and are systematically shifted with respect to the null path counterpart, $P(U_{PS}^i | n_a = 0)$, toward lower energies. The magnitude of the shift is conformation-dependent, although there is no obvious pattern to this dependence. On average, the magnitude of the shift is approximately 3.5 kcal/mol. In light of our definition of the null path $n_a = 0$, the magnitude of the shift is a measure of the work done to add one or two water molecules to the primary hydration shell around either N1 or O2, respectively. The implication is clear: if the primary hydration shells around the peptide group donor and acceptor atoms are fully occupied, changing the number of water molecules in the minimal path has no effect on the potential energy of peptide−solvent interactions. In other words, the central determinant of the thermodynamics of dipeptide

(59) Richards, F. M.; Lim, W. A. *Quart. Rev. Biophys.* **1994**, *26*, 423−498.
(60) Creamer, T. P.; Campbell, M. N. *Adv. Protein Chem.* **2002**, *62*, 263−282.
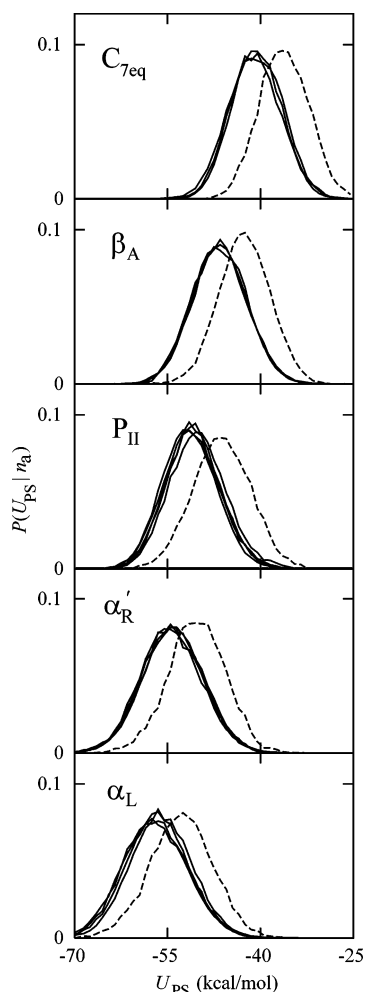
**Figure 8.** Probability distributions of peptide−solvent potential energies for different values of the minimal path $n_a$. Solid curves are for $n_a > 0$; dashed curves are for the null path, $n_a = 0$.

solvation, as viewed from the vantage point of peptide donor and acceptor sites, is the work done to fully hydrate these sites and does not depend on the nature of hydrogen-bonded bridges.

At first glance, the above results contradict the observation that the average peptide−solvent energy and minimal path correlate with each other. However, the observed correlation between these two quantities does not necessarily mean that the preference for large paths facilitates sampling of low-energy peptide−solvent configurations. The correlation exists because both quantities depend on peptide conformation. Indeed, as we have shown, the average minimal path is proportional to the distance between the carbonyl oxygen and amide hydrogen atoms (see Figure 6). The average peptide−solvent energy is also dictated by the degree of overlap between primary hydration shells (Figure 1) around the dipeptide and therefore must be a function of the same distance. Compact conformers that are characterized by lower values of solvent-accessible surface area have larger distances between the peptide donor and acceptor atoms and therefore are more favorably solvated.

**3.4. What Is the Determinant of the Preference for P$_{II}$-like Conformers?** As shown in the previous section, the preference of aqueous alanine dipeptide for P$_{II}$-like conformations does not result from favorable direct interactions of the peptide with surrounding solvent molecules. Also, this prefer-

ence is not a direct consequence of intrapeptide (gas-phase) interactions, which favor conformers such as C$_{7eq}$ that have an intradipeptide hydrogen bond (see Table 2). However, the total potential energy, which is the sum of the two above-mentioned terms, $\Delta U_i = \Delta U_P{}^i + \langle \Delta U_{PS}{}^i \rangle$, gives us P$_{II}$ as the preferred conformation (Table 2). These observations bring us back to the question posed in the Introduction: What interactions are responsible for the preference of alanine dipeptide for P$_{II}$-like conformers in water?

To answer this question, we ascertain if there is a potential energy surface for which P$_{II}$ is the global minimum. On the manifold of fixed bonds and angles, the relative gas-phase potential energy for conformer $i$ is a sum of three different terms, $\Delta U_P{}^i = \Delta U_{vdw}{}^i + \Delta U_{el}{}^i + \Delta U_{torsion}{}^i$, where $\Delta U_{vdw}{}^i$ and $\Delta U_{el}{}^i$ denote contributions due to van der Waals and electrostatic interactions between nonbonded pairs of atoms and $\Delta U_{torsion}{}^i$ is a torsional potential. From the analysis of these three terms, we find that conformers which lie within the favorably populated $\beta$-region on the hydrated free energy surface (Figure 2) are also preferred conformers on the energy surface defined by the sum of van der Waals and torsional potential functions $\Delta U_{vdw}{}^i + \Delta U_{torsion}{}^i$ (Figure 9). Specifically, the P$_{II}$ conformer that is the global minimum on the hydrated free energy surface is close to the global minimum on the van der Waals plus torsional potential energy surface. We take this agreement between the global minima to mean that the origin of the preference for P$_{II}$-like conformers is similar in both cases.

Conformational preferences on the van der Waals plus torsional potential energy surface are mainly determined by minimization of intradipeptide steric conflicts. The latter leads to P$_{II}$ as the dominant conformation. In the total gas-phase potential this preference is masked by attractive electrostatic interactions, which favor conformers such as C$_{7eq}$. In aqueous solution, peptide−solvent interactions counterbalance the effect of intradipeptide attractive interactions and thus unmask underlying conformational preferences that are a consequence of minimizing intradipeptide steric conflicts. This compensatory effect of solvent, rather than direct favorable interactions between the peptide and water molecules, leads to the preference of alanine dipeptide for P$_{II}$-like conformers.

## 4. Discussion

We studied the origin of conformational preferences of alanine dipeptide in water by carrying out Monte Carlo simulations of the aqueous dipeptide in 16 conformations. The conformations were chosen to adequately represent all populated (preferred) regions of the Ramachandran plot. Principal findings of this work are as follows:

(i) The force field of choice—all-atom OPLS with TIP5P water—allows us to recapitulate the preference of alanine dipeptide for the $\beta$-region of ($\phi$, $\psi$) space in general and P$_{II}$-like conformations in particular. Similar results were also obtained by other researchers using different force fields and water models.[19,20,28,30] On the basis of this agreement we conclude that the observed preference for P$_{II}$-like conformations is a robust feature for the alanine dipeptide in liquid water.

(ii) The strength of peptide−solvent interactions is insensitive to the number of hydrogen-bonded water molecules connecting the peptide donor and acceptor atoms. Once the primary hydration shells (Figure 1) around these atoms are fully
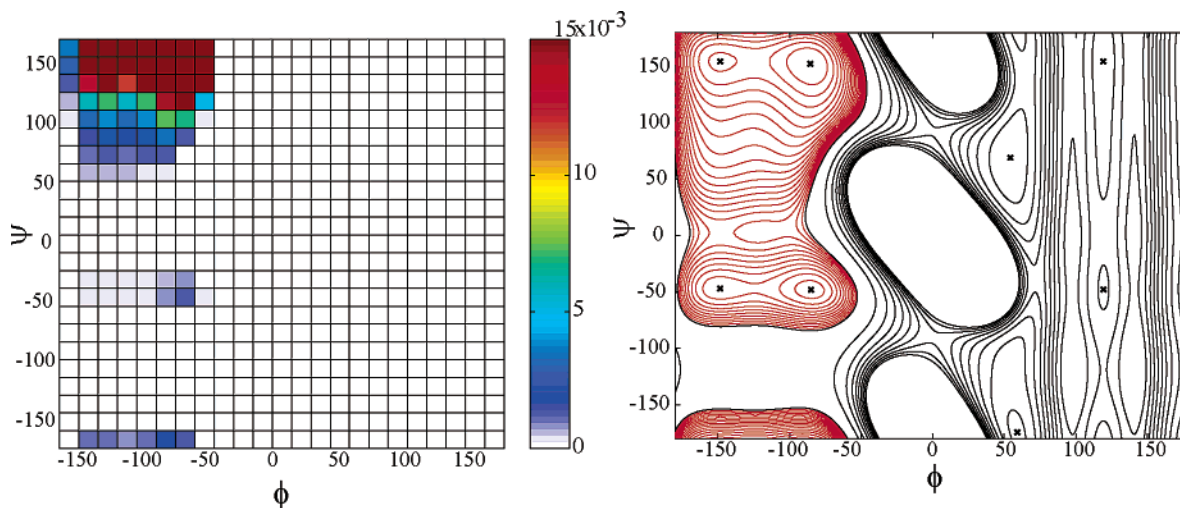
**Figure 9.** Two-dimensional probability density plot [$P(\phi, \psi)$, left panel] and potential energy contour map (right panel) obtained for the alanine dipeptide in terms of the intrapeptide van der Waals and torsional potential functions. Both maps are generated by sampling $(\phi, \psi)$ values on a uniform $1° \times 1°$ grid with fixed bond lengths and bond angles. The adjacent color bar is used to identify low versus high populated regions. Red contours in the right panel are drawn in 0.3 kcal/mol intervals and the black contours are drawn in 6 kcal/mol intervals. The positions of local minima are marked ($\times$). $(\phi, \psi)$ values of the minima, in ascending order of potential energy, are as follows: $(\phi_1, \psi_1) = (-85.54°, 152.56°)$, $(\phi_2, \psi_2) = (-148.39°, 154.84°)$, $(\phi_3, \psi_3) = (-84.91°, -47.98°)$, $(\phi_4, \psi_4) = (-148.56°, -46.88°)$, $(\phi_5, \psi_5) = (54.21°, 69.29°)$, $(\phi_6, \psi_6) = (59.13°, -174.70°)$, $(\phi_7, \psi_7) = (119.28°, 154.28°)$, $(\phi_8, \psi_8) = (119.35°, -47.83°)$.

occupied—on average, two water molecules around each carbonyl oxygen atom and one water molecule around each amide nitrogen atom—there is no thermodynamic advantage or disadvantage to adding more water molecules to the minimal path. This result becomes obvious only when one studies the distribution of water molecules around aqueous alanine dipeptide. In contrast, studies based on embedding the dipeptide in a cluster of water molecules will necessarily lead to the erroneous conclusion that the preference for $P_{II}$-like conformers is due to the formation of single- and double-water bridges between peptide group donor and acceptor atoms.[9,12]

(iii) The decisive factor in solvation thermodynamics is the work done to fully hydrate the peptide. The conformation dependence of this quantity is mainly determined by differences in the work done to fully hydrate the peptide donor and acceptor sites, which in turn is dictated by the overlap of the primary hydration shells around the peptide donor and acceptor atoms (Figure 1). The smaller the overlap, the lower the energy of peptide—solvent interactions.

(iv) The preference of aqueous alanine dipeptide for $P_{II}$-like conformations is not a direct consequence of favorable peptide—solvent interactions, although within the dominant region of $(\phi, \psi)$ space, favorable peptide—solvent interactions prefer $P_{II}$-like geometries over $\beta$-like conformations. Conversely, over the entire accessible $(\phi, \psi)$ space, peptide—solvent interactions favor "compact" conformers such as $\alpha_L$, $P_4$, and $\alpha'_R$, which provide easier access of water molecules to the peptide donor and acceptor sites. The preference for $P_{II}$ arises due to competition between intrapeptide electrostatics and peptide—solvent interactions, which counterbalance each other in aqueous solution. Similar conclusions have been reached on the basis of free energy calculations for alanine dipeptide in a continuum dielectric.[29,61–63] The compensatory effect of solvent unmasks the underlying conformational preferences that are a consequence of minimizing intradipeptide steric conflicts.[32,40,64]

(61) Hopfinger, A. J. *Macromolecules* **1971**, *4*, 731−737.
(62) Yu, H. A.; Pettitt, B. M.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 2425−2434.
(63) Marrone, T. J.; Gilson, M. K.; McCammon, J. A. *J. Phys. Chem.* **1996**, *100*, 1439−1441.
(64) Brant, D. A.; Miller, W. G.; Flory, P. J. *J. Mol. Biol.* **1967**, *23*, 47−65.