ELSEVIER

Journal of Molecular Structure (Theochem) 631 (2003) 277–290

THEO CHEM

www.elsevier.com/locate/theochem

Exploratory conformational analysis of *N*-acetyl-L-Tryptophan-*N*-methylamide. An ab initio study

María L. Ceci[a], María A. Lopez Verrilli[a], Sandra S. Vallcaneras[a], José A. Bombasaro[a], Ana M. Rodríguez[a], Botond Penke[b,c], Ricardo D. Enriz[a,*]

[a]*Departamento de Química, Facultad de Química, Bioquímica y Farmacia, Universidad Nacional de San Luis, Chacabuco 915, 5700 San Luis, Argentina*

[b]*Department of Medical Chemistry, University of Szeged, Dom ter 8, Szeged 6720, Hungary*

[c]*Protein Chemistry Research Group, Hungarian Academy of Sciences, University of Szeged, Dom ter 8, Szeged 6720, Hungary*

Received 17 December 2002; accepted 31 March 2003

Abstract

The full conformational space of *N*-acetyl-L-tryptophan-*N*-methylamide was explored by ab initio MO computations. On the Ramachandran hypersurface of four independent variables $E = E(\phi, \psi, \chi_1, \chi_2)$, 36 conformers were located instead of the expected $3^4 = 81$ stable structures. The relative stabilities of the various conformers were analyzed in terms of sidechain–backbone interactions. A comparative study amongst theoretical calculations and experimental (NMR and X-ray) results was carried out.

Keywords: *N*-acetyl-L-Tryptophan-*N*-methylamide; Peptide conformation; Ab initio MO study; Ramachandran map of tryptophan residue; Sidechain orientation

1. Introduction

The computational and spectroscopic analysis of peptides, incorporating natural and non-natural amino acid residues with and aromatic sidechain, is the focus of many recent studies.

Tryptophan is one of the essential amino acids the human body cannot manufacture it. It is the least abundant in proteins and plays a fundamental role in membrane proteins. The indole sidechain has both hydrophobic and hydrophilic character, and consequently, it partitions at the hydrophobic–hydrophilic interface in lipid bilayers [1]. In the few membrane proteins where a relatively high-resolution structure has been determined, tryptophans are frequently at the bilayer surface where they are oriented so that the indole N–H is directed toward the hydrophilic environment [2–4]. It has been suggested through fluorescence [5], electrophysiological [6], computational [7] and NMR [8] studies that the indole N–H groups may hydrogen bond to the aqueous interface or directly to the lipid molecules. Tryptophan may indeed play an important role in

* Corresponding author.

E-mail addresses: denriz@unsl.edu.ar (R.D. Enriz), mauge_lau@latinmail.com (M.L. Ceci), malopez@unsl.edu.ar (M.A. Lopez Verrilli), ssvallca@unsl.edu.ar (S.S. Vallcaneras), jbombasa@unsl.edu.ar (J.A. Bombasaro), amrodri@unsl.edu.ar (A.M. Rodríguez), penke@ovrisc.mdche.szote.u-szeged.hu (B. Penke).

0166-1280/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0166-1280(03)00261-6

stabilizing membrane proteins through electrostatic interactions at the lipid bilayer surface. Tryptophan is also the source of serotonin (5-hydroxytryptamine) which is a neurotransmitter and a powerful vasoconstrictor [9].

On the other hand, Trp-containing proteins have often been shown to exhibit multiexponential fluorescence decay kinetics in aqueous solution [10,11]. This observation has been rationalized in terms of the conformational heterogeneity of the Trp residue originating from different rotamer conformations [12,13]. This model proposes that the Trp sidechain may adopt different low-energy conformations, due to rotation about χ_1 and χ_2 torsional angles of the sidechain [14] (Fig. 1), with each conformation displaying a distinct decay time. However, data obtained for tryptophan zwitter-ion suggest that excited-state reaction may also be important [15].

Individual rotamers are expected to decay monoexponentially, with lifetimes dependent on the position of the sidechain relative to the indole ring. Thus, it is clearly important to identify the preferred conformations and determine their relative populations and interconversion frequencies.

In addition to the biological and biomedical implications, these aromatic amino acid residues are also important in understanding the aromatic ring-stacking problem [16,17].

Clearly, a deeper understanding of these topics could be enhanced by explicit knowledge of the quantum mechanical conformational properties of Trp. In the present study we report a comprehensive conformational study of *N*-acetyl-L-tryptophan-N-methylamine (**I**) using ab initio calculations.

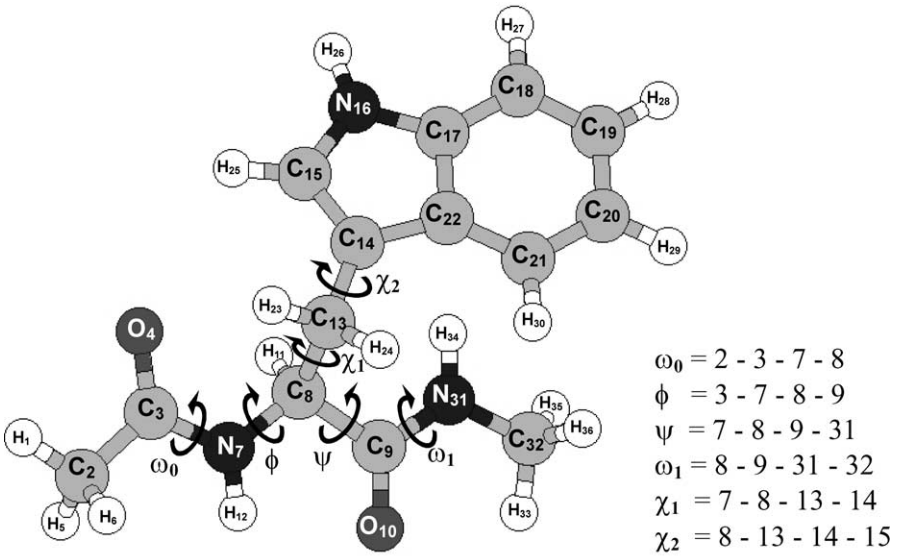
2. Methods

2.1. Conformational analysis

In accordance with the IUPAC–IUB [18] recommendation dihedral or torsional angles were specified [19] within -180 and $180°$ for both backbone (ϕ, ψ) and sidechain (χ_1, χ_2) conformations.

$$-180° \leq \phi \leq 180° \tag{1a}$$

$$-180° \leq \psi \leq 180° \tag{1b}$$



N-acetyl-*L*-tryptophan-**N**-methylamide **(I)**

Fig. 1. Numbering of the atoms and torsional angles for *N*-acetyl-L-tryptophan-*N*-methylamide. The torsional angles are defined in terms of the atoms involved.

$$-180° \leq \chi_1 \leq 180° \quad (1c)$$

$$-180° \leq \chi_2 \leq 180° \quad (1d)$$

According to this definition, when the group, further away from the observer, is rotated clockwise (or the group nearer to the observer rotated counter-clockwise) the dihedral angle is taken to be positive: $0° \rightarrow +180°$. The opposite definition is applicable for negative dihedral angles $0° \rightarrow -180°$.

On the Ramachandran map, the $-180° \leq \phi \leq 180°$ and $-180° \leq \psi \leq 180°$ cut is indicated by the square drawn in broken lines (Fig. 2). But for the graphical presentation of the sidechain conformational potential energy surface (PES), we used the traditional cut ($0° \leq \chi_1 \leq 360°$ and $0° \leq \chi_2 \leq 360°$), similar to that previously suggested by Ramachandran and Sasisekharan [20].

2.2. Molecular computations

Computations were performed on the title compound (**I**) using GAUSSIAN 98 [21] at the RHF/3-21G [22] level of theory. The outcome of the geometry optimizations is presented in Section 3. The total energies are given in hartrees, the relative energies and stabilization energies are given in kcal mol^{-1} using the conversion factor 1 hartree = 627.5095 kcal mol^{-1}. Selected representative conformations were confirmed using RHF/6-31G(d) calculations.

2.3. Stabilization energies

The following isodesmic reaction (2), where R = CH_2-indole, was used to calculate the stabilization energies (3) with respect to the γ_L backbone conformation of N- and C-protected glycine [23,24].

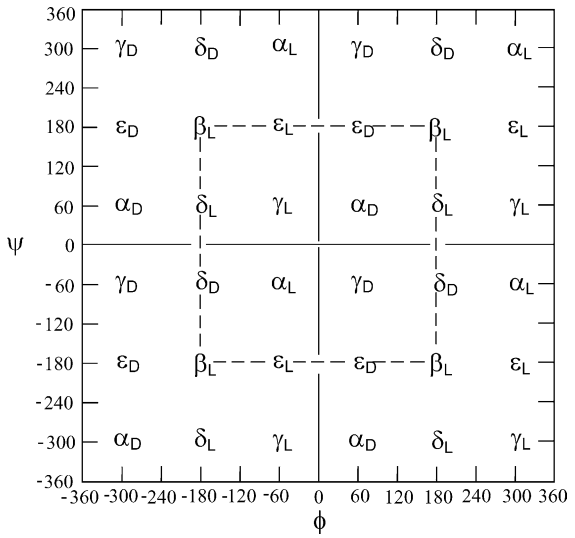


Fig. 2. Topological representation of the Ramachandran map for an N- and C-protected amino acid PCO–NH–CHR–CO–NHQ (P and Q may be H or CH_3) showing two full cycles of rotation: $-360° \leq \phi \leq +360°$; $-360° \leq \psi \leq +360°$. The central box, denoted by broken line, represents the cut suggested by the IUPAC convention. The four quadrants denoted by solid lines are the conventional cuts. Most peptide residues exhibit nine unique conformations labeled as α_D (α_{LEFT}), ε_D, γ_D (C_7^{ax}), δ_L (β_2), β_L (C_5), δ_D (α'), γ_L (C_7^{eq}), ε_L, and α_L (α_{RIGHT}).

$$\underset{\text{reference conformation } \gamma_L}{\text{MeCONH–CH}_2\text{–CONHMe}} + \text{CH}_3\text{–R} \rightarrow \underset{\text{conformation X}}{\text{MeCONH–CHR–CONHMe}} + \text{CH}_3\text{–H} \quad (2)$$

The stabilization energy is defined as follows:

$$\Delta E_{\text{stabilization}} = \{E[\text{MeCONH–CHR–CONHMe}]_X + E[\text{CH}_3\text{–H}]\} - \{E[\text{MeCONH–CH}_2\text{–CONHMe}]_{\gamma_L} + \text{CH}_3\text{–R}]\} \quad (3)$$

This equation is also illustrated graphically below (Eq. (4)) where the molecular structures are symbolizing their total energy values.

$$\Delta E_{\text{stabil.}} = \left[\text{MeOCHN–C(H)(R)–CONHMe} + \text{CH}_3\text{–H} \right] - \left[\text{MeOCHN–C(H)(H)–CONHMe} + \text{CH}_3\text{–R} \right] \quad (4)$$

Table 1
Total energy values of the component molecules for isodesmic reaction computed at RHF/3-21G level of theory

	Molecular system	Energy (hartree)
Me–CONH–CH_2–CONH–Me	γ_L	−451.294243
CH_3–R	R = CH_2–indole	−437.095563
	R = H	−39.9768770

The components' energy values are summarized in Table 1.

3. Results and discussion

3.1. Conformational study

Limiting our considerations only to *trans*-peptide bonds ($\omega_0 \approx \omega_1 \approx 180°$) the full conformational space includes four torsional angles: ϕ, ψ, χ_1 and χ_2 as defined in Fig. 1. Thus, the potential energy hypersurface (PEHS) is function of four independent variables (5).

$$E = E(\phi, \psi, \chi_1, \chi_2) \tag{5}$$

Since we expect three minima (g^+, a, g^-) along each of the variables, multidimensional conformational analysis (MDCA) [25] would dictate the existence of $3^4 = 81$ conformers. These 81 conformers would be distributed evenly, namely 9 sidechain (SC) conformers for each of the 9 backbone (BB) structures. Using MDCA-predicted geometries as input at the HF/3-21G level of theory instead of the expected 81 structures, a total of 36 conformers were located on the PEHS (Eq. (5)). The results of the geometry optimizations are summarized in Table 2 whereas a schematic representation of the 36 existing minima on the PEHS is shown in Fig. 3.

On the basis of the above results, the following observations can be made:

RHF/3-21G calculations predict the existence of 36 conformations for compound **I**, the $\gamma_L(g^+, g^+)$ conformation is being the global minimum. The second higher minimum is the β_L(a, g^-) conformation with 0.93 kcal mol^{-1} above the global minimum. Clearly, in tryptophan, as in all previous cases of L-amino acids studied, conformers with D subscript (α_D, ε_D, γ_D and δ_D) are not preferred due to their relatively high-energy values. These results are typical for most amino acids that have been already studied. The current database, which may provide the basis for comparison includes the following N- and C-protected amino acids, containing a *trans*-peptide bond: Gly [26–33], Ala [26–33], Val [23], Phe [34,35], Ser [36–38], Glu [39,40], Cys [41,42], Gln [43,44] and Ile [45,46]. Preliminary studies have been published on Pro [47], Asp [48] and Sec [49].

It must be pointed out that, although some results about the conformational behavior of Trp is common to others amino acids previously reported, moreover this compound displays conformational intricacies which are clearly 'non-typical'. For example:

(i) The α_L and ε_L conformations, which are usually annihilated using the diamide model, are now energy minima on the Ramachandran PES. Thus, the nine different types of backbone conformations are present in the PEHS of compound **I**. It should be noted that none of the previously reported amino acids posses all the possible BB conformations. Recently, we found the nine possible backbone conformations for Ile [46], however, for this amino acid the only one ε_L conformer possess torsional angles ϕ and ψ very close to those required for the β_L backbone conformation. In addition, we report *N*-acetyl-L-glutamate-*N*-methylamide having α_L and ε_L conformations; however, for this compound the δ_L and δ_D conformations do not represent stable structures.

(ii) None backbone conformations tolerate the torsional angles χ_2 in *anti*. Fig. 3 shows this point very well. This is a striking difference with respect to the rest of amino acids, which will be discussed in detail in Section 3.2.

In order to confirm the above results using more accurate calculations, we perform RHF/6-31G(d) computations. Thus, the global minimum ($\gamma_L(g^+, g^+)$), the second global minimum (β_L(a, g^-)), α_L and ε_L conformations were confirmed using this level of theory.

Table 2
Torsional angles and total energy values for backbone and sidechain conformers of MeCO–Trp–NHMe optimized at RHF/3-21G level of theory. The calculated relative energies (ΔE_{rel}) and stabilization energies (ΔE_{stabil}) are also shown

Final geometry	ϕ (3-7-8-9)	ψ (7-8-9-31)	χ_1 (7-8-13-14)	χ_2 (8-13-14-15)	ω_0 (2-3-7-8)	ω_1 (8-9-31-32)	Energy (Hartree)	ΔE_{rel} (kcal mol^{-1})	ΔE_{stabil} (kcal mol^{-1})
α_D (g^+g^+)	41.33	53.33	49.02	76.70	164.47	−175.87	−848.4130774	10.57	−0.09
α_D (ag^-)	59.61	41.77	−153.21	−78.05	173.78	179.77	−848.4162077	8.60	−2.06
α_D (g^-g^+)	68.95	28.34	−53.50	106.79	170.33	179.28	−848.4167967	8.23	−2.43
	73.56	14.76	−93.29	108.51	172.77	178.29	−848.4142303	9.84	−0.82
α_D (g^-g^-)	65.67	30.12	−56.84	−83.35	172.26	179.34	−848.4177551	7.63	−3.03
ε_D (g^+g^-)	69.61	167.95	119.51	−73.64	−163.08	178.12	−848.4081701	13.65	2.99
	37.00	−134.51	70.02	−3.63	−171.21	−177.00	−848.4057469	15.17	4.51
ε_D (ag^+)	64.68	−174.02	159.58	68.42	−162.48	−179.64	−848.4117719	11.39	0.73
ε_D (ag^-)	65.79	−177.30	−155.56	−116.64	−162.78	−178.20	−848.4123995	10.99	0.23
ε_D (g^-g^+)	73.95	173.32	−54.19	98.85	−165.81	179.05	−848.4127622	10.76	0.11
ε_D (g^-g^-)	65.23	178.64	−60.78	−83.96	−169.22	−179.47	−848.4113665	11.64	0.98
γ_D (g^+g^+)	46.46	−24.86	53.43	81.06	166.82	−179.76	−848.4098201	12.10	1.95
γ_D (g^+g^-)	44.73	−18.98	68.92	−96.47	170.55	−178.32	−848.4105884	12.13	1.47
γ_D (ag^+)	75.11	−62.87	−177.42	76.66	177.06	−178.13	−848.4155994	8.98	−1.68
γ_D (ag^-)	75.54	−58.85	−164.97	−12.54	174.91	−178.49	−848.4208012	5.72	−4.94
	74.35	−68.24	−172.97	−91.93	176.96	178.50	−848.4177769	7.62	−3.04
γ_D (g^-g^+)	79.69	−57.88	−53.08	104.72	171.38	−178.81	−848.4211758	5.48	−5.17
γ_D (g^-g^-)	74.96	−58.64	−61.45	−81.39	171.63	−178.31	−848.4218703	5.05	−5.61
δ_L (g^+g^+)	−127.89	30.24	54.83	83.82	−171.50	176.40	−848.4250188	3.07	−7.59
δ_L (ag^-)	−158.40	54.63	−159.46	−98.93	−175.57	178.15	−848.4207451	5.76	−4.90
δ_L (g^-g^-)	−123.55	26.12	−67.46	−12.32	−173.05	176.33	−848.4195890	6.48	−4.18
β_L (g^+g^+)	−172.30	173.54	54.89	84.68	−178.61	−179.02	−848.4225466	4.62	−603
β_L (g^+g^-)	−164.07	171.34	63.63	−77.85	176.21	177.29	−848.4193971	6.60	−4.06
	−179.95	172.54	46.28	−95.07	−178.62	−178.04	−848.4196776	6.42	−4.23
β_L (ag^+)	−165.87	172.11	−162.03	76.25	177.42	178.17	−848.4237310	3.88	−6.78
β_L (ag^-)	−164.90	171.34	−157.83	−112.47	177.60	179.41	−848.4284366	0.93	−9.73
β_L (g^-g^+)	−126.49	159.48	−66.98	87.10	175.02	176.75	−848.4151395	9.27	−1.39
β_L (g^-g^-)	−128.47	160.86	−73.53	−2.03	173.11	177.05	−848.4214052	5.34	−5.32
δ_D (g^+g^+)	−172.34	−38.08	56.52	91.00	177.25	−178.62	−848.4140864	9.93	−0.73
δ_D (g^+g^-)	160.94	−24.81	59.20	−96.40	173.53	−176.01	−848.4162493	8.58	−2.08
δ_D (ag^+)	193.38	−53.78	−178.35	80.55	174.83	−177.63	−848.4096023	12.75	2.09
δ_D (g^-g^-)	−121.57	−66.41	−71.04	−6.98	175.88	179.06	−848.4103768	12.26	1.60
γ_L (g^+g^+)	−84.83	61.97	46.17	78.57	−174.53	179.96	−848.4299156	0.00	−10.66
γ_L (g^+g^-)	−84.02	60.60	26.30	−105.25	−174.08	179.86	−848.4259364	2.50	−8.16
	−116.40	16.05	48.09	−98.30	−172.40	177.44	−848.4227745	4.48	−6.18
γ_L (ag^+)	−86.79	77.00	−155.52	95.22	−177.05	−146.60	−848.4234323	4.07	−6.59
γ_L (ag^-)	−86.20	72.10	−167.59	−82.08	−175.47	−177.91	−848.4262337	2.31	−8.35
γ_L (g^-g^+)	−85.61	69.25	−49.04	116.33	−170.56	−178.63	−848.4228420	4.44	−6.22
	−115.81	18.72	−63.26	104.03	−167.83	176.45	−848.4156735	8.94	−1.72
γ_L (g^-g^-)	−86.45	68.36	−52.19	−79.75	−171.74	−178.63	−848.4231312	4.26	−6.40
ε_L (g^+g^+)	−72.01	157.74	52.47	26.30	174.45	178.81	−848.4186612	7.06	−3.60
α_L (ag^-)	−67.04	−35.38	−164.23	−20.38	−173.07	−178.79	−848.4148187	9.47	−1.19

The global minimum corresponds to γ_L (g^+g^+) conformation having −848.4299156 hartree total energy. This value is taken as reference value, corresponding to relative energy 0.00 kcal mol^{-1}.

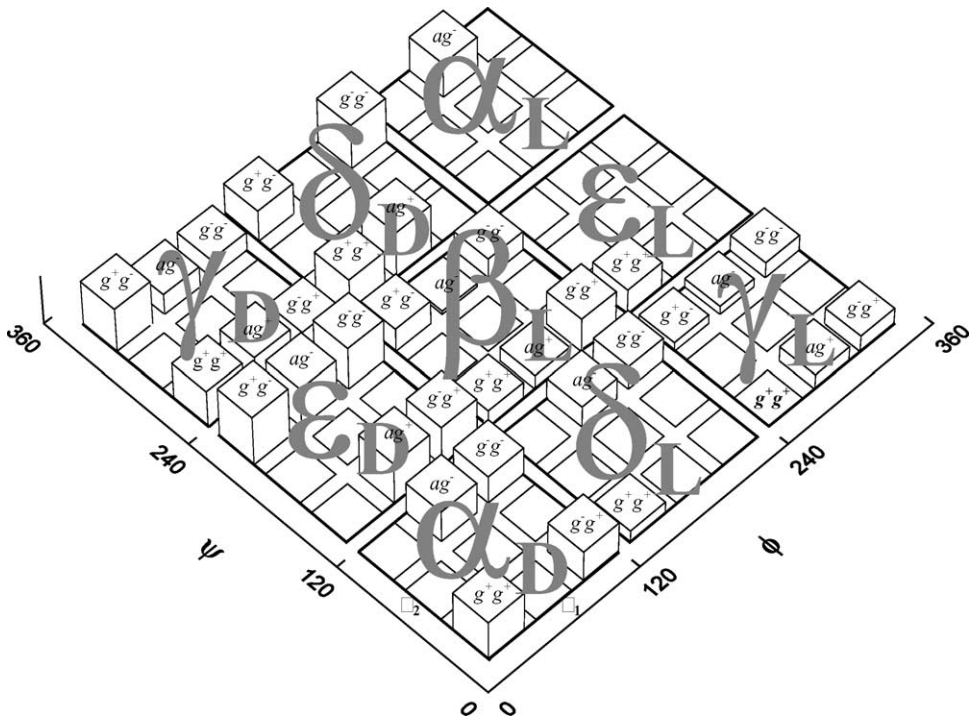


Fig. 3. A schematic representation of the 36 existing minima on the PEHS of four independent variables: $E = E(\phi, \psi, \chi_1, \chi_2)$. The global minimum is denoted in bold and the energy gap above the global minimum is in relationship with the highest of the block.

3.2. Relative energies and sidechain backbone interactions

The six types of backbone–backbone (BB/BB) and sidechain–backbone (SC/BB) interactions found for compound **I** are presented in Fig. 4.

It is difficult to partition the total energy and classify such parts in 'stabilizing' and 'destabilizing'. However, a great deal can be learned by looking beyond the BB/BB and SC/BB interactions. The criteria for the existence of such stabilizing interactions may come from geometry. Any distance shorter than the sum of Van der Waals radii (Table 3) may be taken as diagnostic for such stabilizing interaction. Of the six types of interactions (Fig. 4), four were of the BB/BB type (types 1–4) and two were of the SC/BB category (types 5 and 6). All of them were of the hydrogen-bonding type. The only one conformation possessing type 6 interaction was $\varepsilon_D\,(g^+, g^-)$. It is interesting to note that this form has a distance of 3.10 Å and an angle of 65.95° suggesting a weak hydrogen bonding.

Clearly, not all interactions are contributing to the same extent to the stability of a given conformation; in addition, the sidechain has the ability not only to stabilize but to destabilize as well. An overall summary of intramolecular interactions in compound **I** is given in Table 3. Two representative conformations are shown in Fig. 5.

Sidechain folding is not only interesting but also important for two reasons. On the one hand, sidechain orientation can influence backbone folding via SC/BB interaction. On the other hand, sidechain folding can limit the biological function of the amino acid.

Observing from the results obtained, as shown in Fig. 3, it is clear that compound **I** does not tolerate the *aa* sidechain conformation. To further understand these results, we evaluate the Potential Energy Curves (PEC) of **I** as a function of χ_1 as well as that of χ_2 torsional angles. It should be noted that for compound **I** the overall sidechain orientations are described by the dihedral angles χ_1 and χ_2, respectively; whereas the orientations of the alkyl chain alone are described by the dihedral

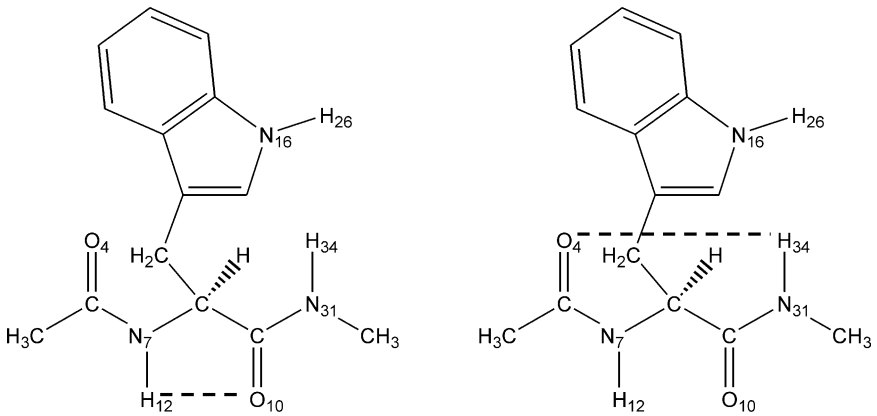


Fig. 4. Various types of intramolecular interactions (i.e. BB/BB and SC/BB) in *N*-acetyl-L-tryptophan-*N*-methylamide.

Table 3
Summary of intramolecular interactions in *N*-acetyl-L-Tryptophan-*N*-methylamine optimized at RHF/3-21G level of theory

Conformation	ΔE_{rel} (kcal mol^{-1})	Interaction type[a]	Distance (Å)[a]	Angle (°)[a]
α_D (g^+, g^+)	10.57	4	2.53	99.99
α_D (a, g^-)	8.60	4	2.38	101.93
α_D (g^-, g^+)	8.23	4	2.31	105.04
α_D (g^-, g^+)	9.84	2	2.99	99.26
		4	2.31	106.27
α_D (g^-, g^-)	7.63	4	2.32	104.61
ε_D (g^+, g^-)	15.17	5	3.10	65.95
ε_D (a, g^+)	11.39	6	2.37	138.16
γ_D (g^+, g^+)	12.10	2	1.73	156.36
		4	2.70	99.15
γ_D (g^+, g^-)	12.13	2	1.76	156.27
		4	2.68	98.97
γ_D (a, g^+)	8.98	2	1.93	146.86
		4	2.77	92.99
γ_D (a, g^-)	5.72	2	1.91	148.64
		4	2.73	93.53
γ_D (a, g^-)	7.62	2	1.94	145.71
		4	2.81	92.86
γ_D (g^-, g^+)	5.48	2	1.95	149.77
		4	2.70	94.25
γ_D (g^-, g^-)	5.05	2	1.89	149.66
		4	2.73	94.57
δ_L (g^+, g^+)	3.07	4	2.31	104.23
δ_L (a, g^-)	5.76	4	2.55	94.22
δ_L (g^-, g^-)	6.48	4	2.30	104.36
β_L (g^+, g^+)	4.62	1	2.10	109.10
β_L (g^+, g^-)	6.60	1	2.09	109.03
β_L (g^+, g^-)	6.42	1	2.06	110.83
β_L (a, g^+)	3.88	1	2.08	109.35
β_L (a, g^-)	0.93	1	2.08	109.35
β_L (g^-, g^+)	9.27	1	2.29	100.92
β_L (g^-, g^-)	5.34	1	2.26	101.56
δ_D (g^+, g^+)	9.93	4	2.35	102.17
		3	2.36	101.99
δ_D (g^+, g^-)	8.58	4	2.20	106.77
		3	2.43	92.54
δ_D (a, g^+)	12.75	4	2.57	96.88
		3	2.54	98.50
δ_D (g^-, g^-)	12.26	4	2.78	90.97
		3	2.72	94.75
γ_L (g^+, g^+)	0.00	2	1.98	147.47
		4	2.71	92.63
γ_L (g^+, g^-)	2.50	2	1.95	148.28
		4	2.69	92.95
γ_L (g^+, g^-)	4.48	4	2.27	106.61
γ_L (a, g^+)	4.07	2	2.10	138.15
		4	2.86	88.05
γ_L (a, g^-)	2.31	2	2.04	141.61
		4	2.81	89.70
γ_L (g^-, g^+)	4.44	2	1.97	144.58
		4	2.78	91.42
γ_L (g^-, g^+)	8.94	4	2.12	108.45
γ_L (g^-, g^-)	4.26	2	1.99	144.70
		4	2.77	91.33
α_L (a, g^-)	9.47	4	2.33	103.84

[a] Definition is given in Fig.

angle χ_1, while the orientations of the ring system alone are described by χ_2. These curves, computated for the γ_L backbone conformation, are presented in Fig. 6.

The PEC obtained for χ_1 (Fig. 6, top) exhibit three minima; the first minimum is at a dihedral angle of about 20° and the others at approximately −170 and −60°, respectively. The orientations of $-CH_2$ group at the dihedrals 20 and −60° refer to *gauche* + and *gauche* − orientations and −170° to anti orientation. This curve displays a relatively shallow low-energy region between −170 and 20° and a large barrier between 20 and −170°. The rotational barrier at −30° is very low (1.04 kcal mol^{-1}), at −120° is 3.82 kcal mol^{-1}, whereas the barrier at 120° is 20.18 kcal mol^{-1}.

Looking at the torsional angle χ_2 (Fig. 6, bottom) we see that the global minimum was found to be the one in which the ring system was anti-perpendicular to the C_8–C_{13} bond ($\chi_2 \approx -90°$). Of course, this minimum occurs twice in full 360° rotation due to the fact that the indole ring is planar. However, the local minimum (at $\chi_2 \approx 90°$) possess 5.60 kcal mol^{-1} above the global minimum. All in all, 3 minima (g^+, a, g^-) along χ_1 and 2 minima (g^+, g^-) along χ_2 suggest a grand total of $3 \times 2 = 6$ sidechain orientation for the γ_L backbone conformation. This is exactly what has been found by geometry optimization as shown in Table 2. The barrier height between the anti-perpendicular χ_2 and perpendicular χ_2 forms is 6.21 kcal mol^{-1}. It is interesting to note that an extremely large barrier was obtained for χ_2 at approximately 150° (47.21 kcal mol^{-1}). This result

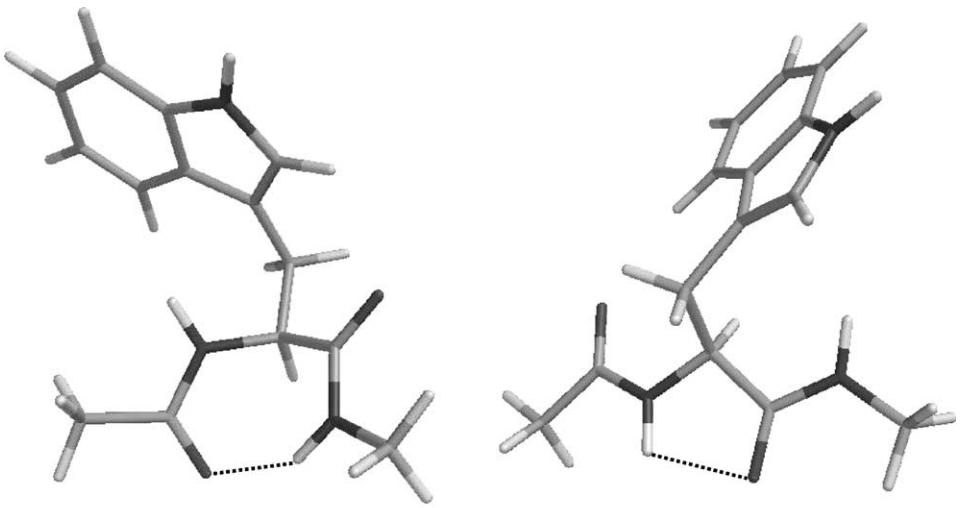


Fig. 5. Spatial view of the γ_L (g^+, g^+) and β_L (a, g^-) conformations obtained from RHF/3-21G calculations. The hydrogen bonds stabilizing the respective conformations are shown in this figure.

account for why compound **I** does not tolerate anti conformations for χ_2 (Fig. 3).

The relative barrier heights obtained for χ_1 and χ_2 would imply that for *N*-acetyl-L-tryptophan-*N*-methylamide, χ_1 inter-conversions would occur more frequently than χ_2 inter-conversions. RHF/3-21G computations predict that barriers ranging from 1.04 to 6.21 kcal mol^{-1} are separating the different conformations (relative to the sidechain orientation) and therefore the conformational inter-conversions are somewhat restricted but still available for this compound.

These results are in qualitative agreement with calculations previously reported for tryptophan [14,50–52]. NMR studies [53–56] and surveys of the conformations of tryptophan residues in proteins (from X-ray structural data) [57,58] identify similar preferred rotamers.

3.3. Stabilization energies

Stabilization energy is a measure of the stabilization ($\Delta E_{stabil.} < 0$) or destabilization ($\Delta E_{stabil.} > 0$) exerted by the sidechain on the backbone with respect to hydrogen, i.e. the sidechain of glycine. The stabilization energy is calculated according to Eqs. (3) and (4) which are based on the corresponding isodesmic reaction (2). Traditionally, the global minimum is used for such calculation and also in the case of peptides. It was an obvious choice to use the γ_L BB conformation.

The ($\Delta E_{stabil.}$ (γ_L) values, summarized in Table 2, are presented graphically in Fig. 7. An interesting pattern is emerging with respect to the role of the sidechain orientation in the stabilizing or destabilizing process. For example γ_L, β_L and α_D backbones are stabilized by all sidechain orientations. In the only one conformation obtained for α_L and ε_L forms, the sidechain orientation is stabilizing the respective structure. In contrast, all sidechain orientations exert only a destabilizing effect on the ε_D conformations. The γ_D and δ_D backbone conformations have a mixture of stabilizing and destabilizing sidechain orientations.

A comparison to other amino acids is shown in Fig. 8. Tryptophan (Trp) falls between delta-alanine (ΔAla) and asparagine (Asn).

3.4. Correlation between natural occurrence of conformers and computer stability

The validity of calculations reported here may be assessed by comparing the predicted structures with those derived experimentally, either by

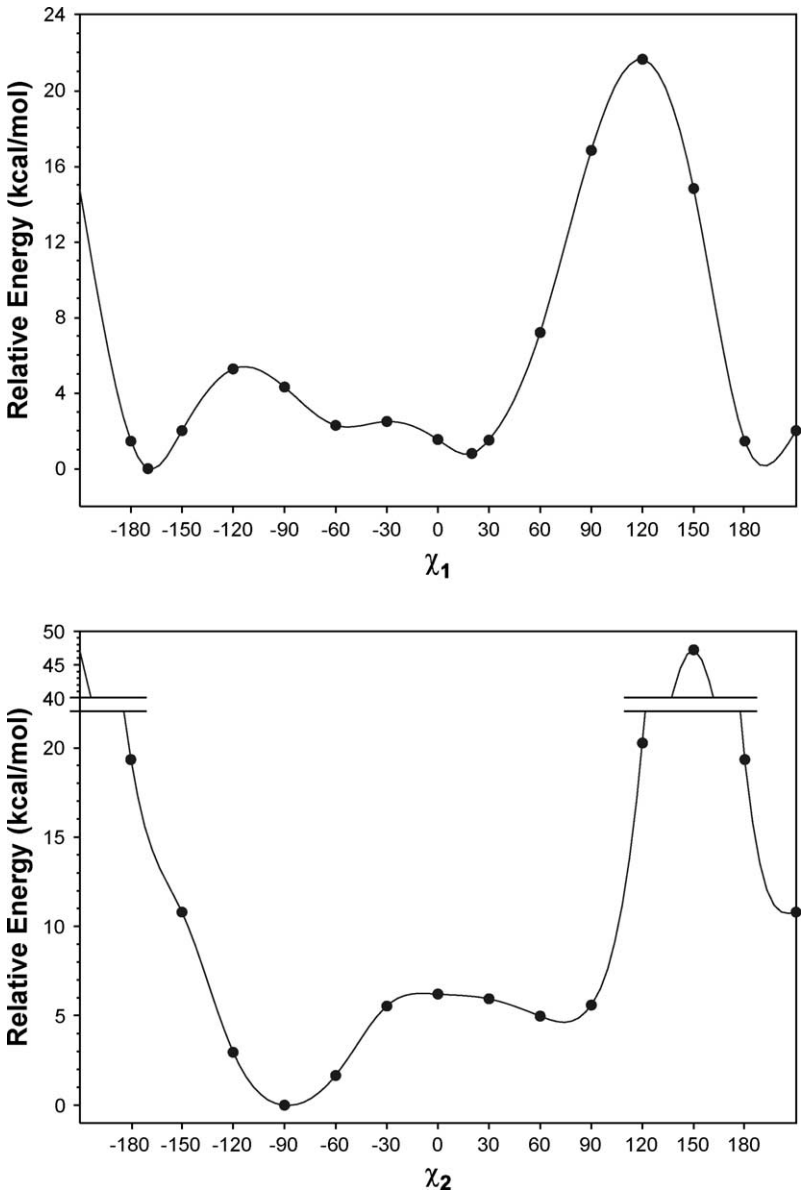


Fig. 6. Potential energy curves of *N*-acetyl-L-tryptophan-*N*-methylamide in its γ_L backbone conformation from RHF/3-21G relaxed single scan computations. Upper curve: $E = E(\chi_1)$, χ_1 measures the rotation about the C_8-C_{13} bond keeping $\chi_2 = -90°$. Lower curve: $E = E(\chi_2)$, χ_2 measures the rotation about the $C_{13}-C_{14}$ bond keeping $\chi_1 = 180°$.

X-rays crystallographic or solution studies (NMR). Thus, the comparison of relative energies obtained from ab initio calculations and relative probabilities of conformers using a non-homologous database is a possibility for this cross-validation.

Using a recent (February 2002) X-ray and NMR-determined protein data set of non-homologous proteins [59], a population-distribution map was generated. The backbone conformers of 1347 Trp residues, found in a total of 331 proteins, were plotted,

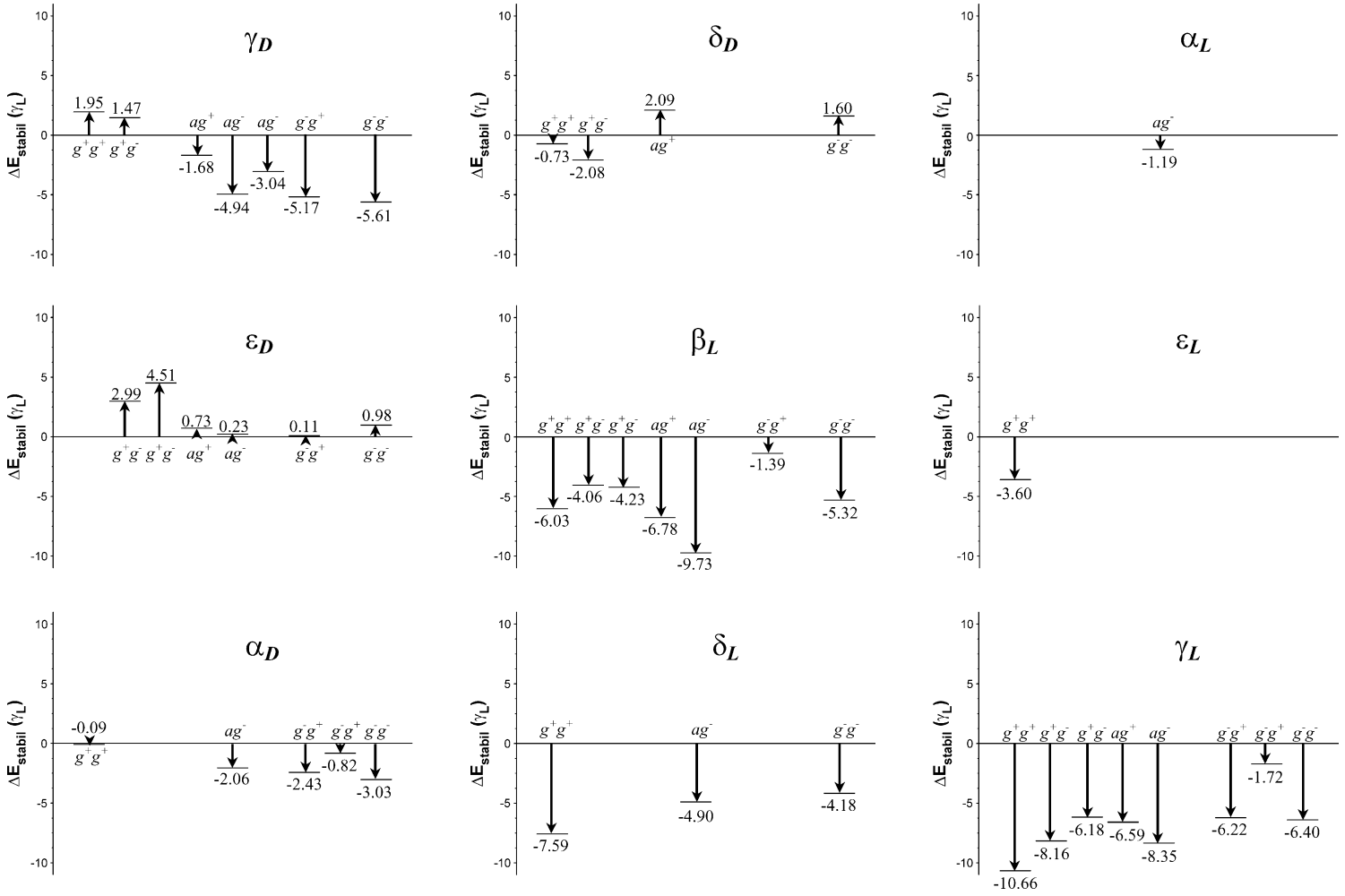


Fig. 7. A graphical presentation of the $\Delta E_{\text{stabil.}}$ (γ_L) values for all existing backbone and sidechain conformations of *N*-acetyl-L-tryptophan-*N*-methylamide.

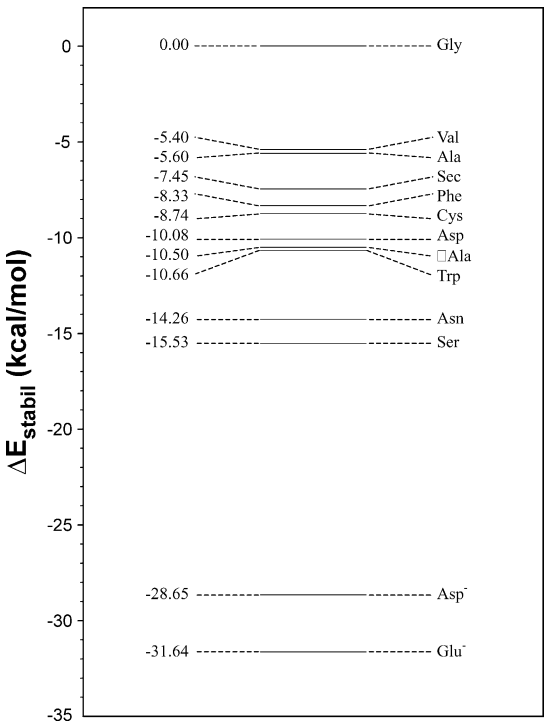


Fig. 8. Relative stabilization energies of various sidechain (R) of N- and C-protected amino acids: PCONH–CHR–CONHQ (where P and Q may be H or Me).

showing ϕ vs. ψ values (Fig. 9, top). The overall appearance of the Ramachandran surfaces is similar to other previously reported Ramachandran maps (Glu [40], Cys [42], Phe [60] and Ser [61]).

To perform a comparison between calculated and observed backbone conformers, an additional plot was made with the RHF/3-21G results (Fig. 9, bottom).

Comparing these data a promising overall similarity emerged. The experimental (X-ray and NMR-) data indicate three highly populated zones, the first one corresponds to the α_L (right-hand α-helix) and δ_D regions; the second is the β_L (extended β-strand), γ_L (inverse-γ-turn), δ_L and ε_L regions and the third is the α_D zone, which corresponds to the left-hand α-helix region (the three zones are denoted by dotted lines in Fig. 9). It is interesting to note that RHF/3-21G calculations indicate these zones as the preferred conformations of **I**. Only four conformations included in a range of 10 kcal mol^{-1} with respect to the global minimum were not found in the experimental data. These conformations are γ_D and possess 5.05, 5.48, 7.62 and 8.98 kcal mol^{-1} above the global minimum, therefore it is clear that they are not preferred forms for **I**. These conformations are denoted with empty circles in Fig. 9, bottom. Such a correlation permit us

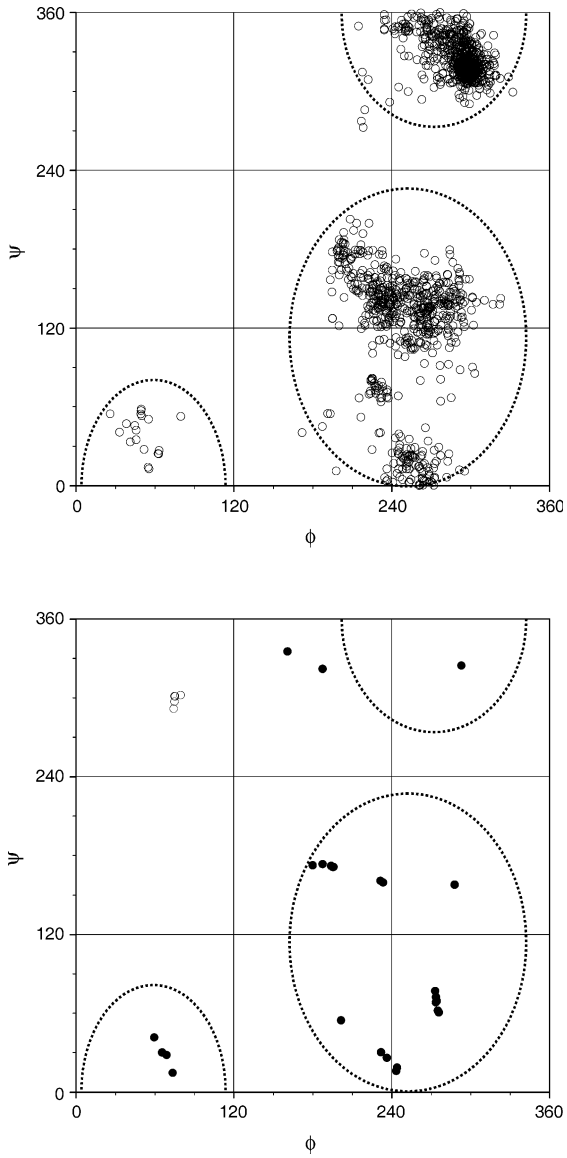


Fig. 9. Top: backbone conformers of all 1347 Trp residues taken from 331 non-homologous proteins. Using their backbone dihedral parameters all the above Trp residues were plotted on a [ϕ, ψ] map. Bottom: locations of calculated ab initio RHF/3-21G of MeCO–Trp–NHMe backbone conformers on a [ϕ, ψ] map. Only conformations with less than 10 kcal mol^{-1} above the global minimum were plotted.

to assume that if the diamide model is relevant to the description of main-chain folding of proteins, then the most stable conformers should have the lowest energy.

4. Conclusions

We have presented a computational study of the conformational preferences of *N*-acetyl-L-tryptophan-*N*-methylamide using ab initio calculations.

A total of 36 minimum energy conformations were characterized at the RHF/3-21G level. The lowest energy structure corresponds to a γ_L backbone orientations with g^+, g^+ conformation in the side-chain. Ab initio calculations predict the γ_L and β_L conformations as the preferred forms of the Trp. These results are in agreement with those previously reported for several amino acids. However, Trp displays some conformationally atypical behaviors with respect to the rest amino acids previously reported: (i) It has the nine backbone conformations in the PEHS; it is the first single amino acid possessing all the backbone conformations. (ii) None backbone conformations tolerate the torsional angle χ_2 in anti. This result might be attributed to the significant steric hindrance due to the indole ring.

It has been previously reported that a polar sidechain produce a significant influence in the backbone conformations of different amino acids like Ser, Cys, Phe and Gln. On the bases of our results, this effect is particularly apparent in Trp as well.

Acknowledgements

This work was supported by grants from Fundación Antorchas and Universidad Nacional de San Luis (UNSL)-Argentina. R.D. Enriz is member of career researcher of CONICET-Argentina.

References

[1] R.E. Jacobs, S.H. White, Biochemistry 28 (1989) 3421.
[2] M. Michel, J. Deisenhofer, Curr. Top. Membr. Trans. 36 (1990) 53.
[3] R. Henderson, J.M. Baldwin, T.A. Ceska, F. Zemblin, E. Beckmann, K.H. Downing, J. Mol. Biol. 213 (1990) 899.
[4] A. Chattopadhyay, M.G. Mc Namee, Biochemistry 30 (1991) 7159.
[5] S.F. Scarlatta, Biochemistry 30 (1991) 9853.
[6] A.M. O' Connell, R.E. II Koeppe, O.S. Andersen, Science 250 (1990) 1256.
[7] G.H.W. Meulendijks, T. Sonderkamp, J.E. Dubois, R.J. Nielsen, J.A. Kremers, H.M. Buck, Biochem. Biophys. Acta 979 (1989) 321.
[8] N.D. Lazo, W. Hu, T.A. Cross, J. Chem. Soc. Chem. Commun. (1992) 1529.
[9] L.F. Pisterzi, D.R.P. Almeida, G.A. Chass, L.L. Torday, J.Gy. Papp, A. Varro, I.G. Csizmadia, Chem. Phys. Lett. 365 (2002) 542.
[10] J.M. Beechem, L. Brand, Annu. Rev. Biochem. 54 (1985) 43.
[11] M. Eftink, in: C.H. Suelter (Ed.), Methods of Biochemical Analysis: Protein Structure Determination, vol. 35, Wiley, New York, 1991, p. 129.
[12] J.B.A. Ross, H.R. Wyssbrod, R.A. Porter, G.P. Schwartz, C.A. Michaels, W.R. Laws, Biochemistry 31 (1992) 1585.
[13] K.J. Willis, A.G. Szabo, Biochemistry 31 (1992) 8924.
[14] H.L. Gordon, H.C. Jarrell, A.G. Szabo, K.J. Willis, R.L. Somorjai, J. Phys. Chem 96 (1992) 1915.
[15] K.J. Willis, A.G. Szabo, D.T. Krajcarski, Chem. Phys. Lett. 182 (1991) 614.
[16] O. Yamauchi, Pure Appl. Chem. 67 (1995) 297.
[17] S. Ghalem, B.T. Fan, L. Xiao, G. Lapluye, J. Chem. Phys.—Chem. Biol. 91 (1994) 113.
[18] IUPAC-IUB Comission on Biochemical Nomenclature, Biochemistry 9 (1970) 3471.
[19] O. Farkas, A. Perczel, J.F. Marcoccia, M. Hollosi, I.G. Csizmadia, J. Mol. Struct. (Theochem) 331 (1995) 27.
[20] I. Ramachandran, V. Sasisekharan, Adv. Protein Chem. 23 (1968) 283.
[21] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople, Gaussian 98, Revision A.7, Gaussian, Inc., Pittsburgh PA, 1998
[22] J.S. Binkley, J.A. Pople, W.J. Hehre, J. Am. Chem. Soc. 102 (1980) 939.
[23] W. Viviani, J.-L. Rivail, A. Perczel, I.G. Csizmadia, J. Am. Chem. Soc. 115 (1993) 8321.
[24] M.A. McAllister, G. Endredi, W. Viviani, A. Perczel, P. Csaszar, J. Ladik, J.-L. Rivail, I.G. Csizmadia, Can. J. Chem. 73 (1995) 563.

[25] M.R. Peterson, I.G. Csizmadia, Progr. Theor. Org. Chem. 3 (1982) 190. I.G. Csizmadia, I.G. Csizmadia,New Theorical Concept for Understanding Organic ReactionsReidel (1989). A. Perczel, J.G. Angyan, M. Kastár, W. Viviani, J.L. Rivail, J.F. Marcoccia, I.G. Csizmadia, J. Am. Chem. Soc. 113 (1991) 6256.
[26] T. Head-Gordon, M. Head-Gordon, M.J. Frisch, C.L. Brooks, J.A. Pople, J. Am. Chem. Soc. 113 (1991) 5989.
[27] A. Perczel, J.G. Angyan, M. Rajtar, W. Viviani, J.L. Rivail, J.F. Marcoccia, I.G. Csizmadia, J. Am. Chem. Soc. 113 (1991) 6256.
[28] M.A. McAllister, A. Perczel, P. Csaszar, W. Viviani, J.L. Rivail, I.G. Csizmadia, J. Mol. Struct. (Theochem) 288 (1993) 161.
[29] S.J. Weiner, V.C. Singh, T.J. O'Donnell, P.A. Kollman, J. Am. Chem. Soc. 106 (1984) 6243.
[30] A.M. Sapse, L.M. Fugler, D. Cowburn, Int. J. Quantum Chem. 29 (1986) 1241.
[31] T.C. Cheam, S. Krimm, J. Mol. Struct. (Theochem) 188 (1989) 15.
[32] T.C. Cheam, S. Krimm, J. Mol. Struct. (Theochem) 193 (1989) 1.
[33] T. Head-Gordon, M. Head-Gordon, M.J. Frish, C. Brooks, J. Pople, Int. J. Quantum Chem. Quantum Biol. Symp. 16 (1989) 311.
[34] O. Farkas, M.A. McAllister, J.H. Ma, A. Perczel, M. Hollósi, I.G. Csizmadia, J. Mol. Struct. (Theochem) 369 (1996) 105.
[35] A. Perczel, O. Farkas, I.G. Csizmadia, Can. J. Chem. 75 (1997) 1120.
[36] O. Farkas, A. Perczel, J.F. Marcoccia, M. Hollósi, I.G. Csizmadia, J. Mol. Struct. (Theochem) 331 (1995) 27.
[37] A. Perczel, O. Farkas, I.G. Csizmadia, J. Comput. Chem. 17 (1996) 821.
[38] A. Perczel, O. Farkas, I.G. Csizmadia, J. Am. Chem. Soc. 118 (1996) 7809.
[39] M.F. Masman, M.G. Amaya, A.M. Rodriguez, F.D. Suvire, G.A. Chasse, O. Farkas, A. Perczel, R.D. Enriz, J. Mol. Struct. (Theochem) 543 (2001) 203.
[40] M.F. Massman, M.A. Zamora, A.M. Rodriguez, N.G. Fidanza, N.M. Peruchena, R.D. Enriz, I.G. Csizmadia, Eur. Phys. J. D. 20 (2002) 531.
[41] M.A. Zamora, H.A. Baldoni, J.A. Bombasaro, M.L. Mak, A. Perczel, O. Frakas, R.D. Enriz, J. Mol. Struct. (Theochem) 540 (2001) 271.
[42] M.A. Zamora, H.A. Baldoni, A.M. Rodriguez, R.D. Enriz, C.P. Sosa, A. Perczel, A. Kucsman, O. Farkas, E. Deretey, J.C. Vank, I.G. Csizmadia, Can. J. Chem. 80 (2002) 832.
[43] A.M. Tarditi, M.W. Klipfel, A.M. Rodriguez, F.D. Suvire, G.A. Chasse, O. Farkas, A. Perczel, R.D. Enriz, J. Mol. Struct. (Theochem) 545 (2001) 29.
[44] M.W. Klipfel, M.A. Zamora, A.M. Rodriguez, N.G. Fidanza, R.D. Enrriz, I.G. Csizmadia, Exploration of the full conformational space of *N*-acetyl-L-glyamine-*N*-methylamide. An ab initio and DFT study. J. Phys. Chem. A (2003) in press.
[45] M.N. Barroso, E.S. Cerutti, A.M. Rodriguez, E.A. Jauregui, O. Farkas, A. Perczel, R.D. Enriz, J. Mol. Struct. (Theochem) 548 (2001) 21.
[46] F.C. Calaza, M.V. Rigo, A.N. Rinaldoni, M.F. Masman, A.M. Rodriguez, R.D. Enriz, Comprehensive conformational analysis of *N*-acetyl-L-isoleucine-*N*-methylamide. An ab initio study. J. Mol. Struct. (Theochem) (2003) in press.
[47] H.A. Baldoni, A.M. Rodriguez, G. Zamarbide, R.D. Enriz, Ö. Farkas, P. Csaszar, L.L. Torday, C.P. Sosa, I. Jakli, A. Perczel, M. Hollosi, I.G. Csizmadia, J. Mol. Struct. (Theochem) 465 (1999) 79.
[48] S.J. Salpietro, A. Perczel, Ö. Farkas, R.D. Enriz, I.G. Csizmadia, J. Mol. Struct. (Theochem) 497 (2000) 39.
[49] J.C. Vank, C.P. Sosa, A. Perczel, I.G. Csizmadia, Can. J. Chem. 78 (2000) 395.
[50] R.A. Engh, L.X.Q. Chen, G.R. Fleming, Chem. Phys. Lett. 126 (1986) 365.
[51] S.S. Zimmerman, M.S. Pottle, G. Némethy, H.A. Scheraga, Macromolecules 10 (1977) 1.
[52] J.S. Anderson, G.S. Bowitch, R.L. Brewster, Biopolymers 22 (1983) 2459.
[53] F.R. Cavanaugh, J. Am. Chem. Soc. 92 (1970) 1488.
[54] A. Bacci, V. Rizzo, P. Skrabal, P.L. Luisi, J. Am. Chem. Soc. 101 (1979) 5170.
[55] J. Kobayaski, T. Higashima, S. Sekido, T. Miyazawa, Int. J. Pept. Protein Res. 17 (1981) 486.
[56] B. Dezube, C.M. Dobson, C.E. Teagne, J. Chem. Soc., Perkin Trans. 2 (1981) 730.
[57] J. Janin, S. Wodak, M. Levitt, B. Maigret, J. Mol. Biol. 125 (1978) 357.
[58] T.N. Bhat, V. Sasisekharan, M. Vijayan, Int. J. Pept. Protein. Res. 13 (1979) 170.
[59] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissing, I. Shindyalov, P. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235.
[60] A. Perczel, O. Farkas, I. Jakli, I.G. Csizmadia, J. Mol. Struct. (Theochem) 455 (1998) 315.
[61] I. Jakli, A. Perczel, O. Farkas, C.P. Sosa, I.G. Csizmadia, J. Comput. Chem. 21 (2000) 626.