

Gas phase folding of an (Ala)₄ neutral peptide chain: spectroscopic evidence for the formation of a β -hairpin H-bonding pattern†

Eric Gloaguen,^{ab} Rodolphe Pollet,^{ab} François Piuze,^{ab} Benjamin Tardivel^{ab} and Michel Mons^{*ab}

Received 9th September 2009, Accepted 14th October 2009

First published as an Advance Article on the web 2nd November 2009

DOI: 10.1039/b918670k

IR and UV laser spectroscopy of an Ala-based 4-residue model peptide recorded under gas phase isolated conditions provides evidence for the intrinsic stability of compact folded structures resembling the extremity of a β -hairpin, with a C₁₄ H-bond bridging the two ends of the chain, and enables us to assess the capabilities of new quantum chemistry techniques to account for dispersive interactions in a medium-size molecule.

Protein modelling is of great importance for the understanding of function of biological systems. Yet protein and peptide modelling is still facing difficulties in treating properly at the molecular level specific ubiquitous interactions like dispersive interactions. Dispersion is indeed still challenging for quantum chemistry, despite recent advances in the development of functionals specifically designed to tackle this issue, *e.g.*, by adding explicit semi-empirical terms (DFT-D).^{1–5} In this context, gas phase data on neutral peptides turn out to provide useful benchmarks on key systems for the assessment of theoretical methods. Several studies have demonstrated that typical secondary structures of proteins, like β -turns or ₃₁₀ helices, can be observed as in isolated neutral molecules and precisely characterized using laser optical techniques, including double-resonance spectroscopy.^{6–18} However, apart from a few examples,^{18,19} larger secondary structures, like β -hairpins or α -helices, still remain to be observed on neutral isolated peptides, presumably because they occur for larger sizes, which are more difficult to study both theoretically and experimentally. Indeed, beyond a few residues, the features of UV and IR spectra become difficult to resolve¹⁹ and precise assignments are hampered by the extensive conformational exploration required and the lack of reliability of the synthesized vibrational spectra. We present here original results on a 4-residue Ala-based model peptide, Ac-(Ala)₄-O-Bzl (where Ac- and -O-Bzl stand for *N*-acetyl and benzyloxy groups, respectively), which also exhibits a flexible aromatic tail very sensitive to dispersive interactions. On such a medium-size molecule, using optical spectroscopic methods enables us to determine the intrinsic folding adopted by a

neutral peptide chain under isolated conditions. Emphasis is here put on the folding patterns adopted by the 4-residue backbone including a β -hairpin structure together with the issue of the relevant modelling methods for the structural and vibrational properties of these species.

The molecules were vaporized using a laser desorption procedure described elsewhere.²⁰ Subsequent cooling is achieved through collisions in a supersonic expansion. The resulting molecules are cold and their UV absorption spectrum in the near UV region (lowest $\pi\pi^*$ transition of the benzyl chromophore), as revealed by mass-resolved resonant two-photon ionization (R2PI), is composed of two well-separated series of narrow lines indicating two conformers labeled A and B (Fig. 1a). This spectral selectivity, due to different environments of the benzyl moiety, enables us to carry out IR/UV double-resonance spectroscopy experiments, which lead *in fine* to single-conformation IR spectra. The IR spectra obtained in the NH-stretch region (Fig. 1b) are very sensitive to the local environment of the NH bonds and their analysis permits a direct probe of the H-bonding of the conformer studied. Both conformers A and B exhibit relatively similar IR spectra, with three red-shifted bands in the region of H-bonded NH's (wavenumbers <3440 cm⁻¹);¹⁰ the missing band, corresponding to the fourth NH, being presumably in the free NH region range (\sim 3480–3500 cm⁻¹), not accessible with the present IR source. Whereas the two most red bands that correspond to strong H-bonds differ from one conformer to the other, the third one is observed at 3383 cm⁻¹ in both cases. In addition, the absence of bands in the 3420–3450 cm⁻¹ region indicates the absence of any weak H-bond in the system, like an NH- π bond.

Analysis of the possible H-bonding structures in a backbone of this size, benefiting from our experience on smaller species,¹⁰ suggests that only the four types of H-bonding networks listed below and displayed Fig. 1c can account for 3 H-bond conformations (H-bonding networks will be described according to the type of H-bonds or interactions (C₅, C₇, C₁₀) each NH along the chain is engaged in; *x* standing for a free NH): either triple C₇ structures, noted *x*-7-7-7, or folded backbone structures, in which the first amide interacts with the ester moiety of the C-terminus through a C₁₄ H-bond, giving rise to 14-7-7-*x*, 14-7-*x*-10 or 14-*x*-10-10 networks (Fig. 1c); a fifth 14-*x*-10-7 network is actually forbidden for steric reasons.

For a further insight on the conformational landscape, extensive explorations of the system hypersurface have been carried out, followed, for those structures having 3 or 4

^aCEA, IRAMIS, Service des Photons, Atomes et Molécules, F-91191 Gif-sur-Yvette, France. E-mail: michel.mons@cea.fr; Fax: 33 16908 1213; Tel: 33 16908 2001

^bCNRS, LFP, URA 2453, BioMolecular Structures Group, F-91191 Gif-sur-Yvette, France

† Electronic supplementary information (ESI) available: Appendix S1: conformational exploration procedure; Fig. S1, Table S1: RI-DFT-D calculated optimized structures, structural parameters, energetics and vibrational frequencies of selected conformations. See DOI: 10.1039/b918670k

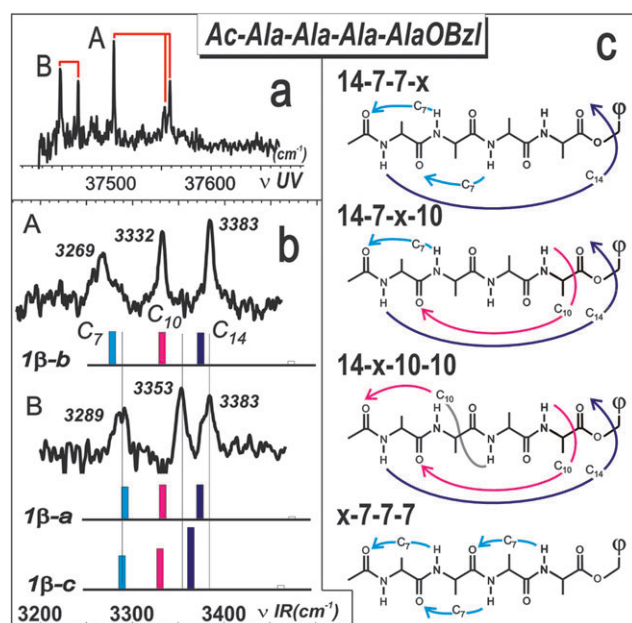


Fig. 1 (a) Near UV spectrum of the Ac-(Ala)₄-O-Bzl molecule as obtained from R2PI, with the spectral signatures of two conformers labeled A and B; (b) IR/UV double resonance spectra of A and B, in the NH stretch region compared with IR calculated stick spectra of the three conformations, which provide the best fit to experiment (see Fig. 2 and Fig. S1†). On the B spectrum the apparent band broadening is assigned to the modest signal-to-noise ratio achieved on this less populated conformer. Synthetic spectra are obtained from scaled (0.9744) RI-B97-D/TZVPP harmonic frequencies; (c) The four bonding schemes compatible with a 3 H-bond network.

H-bonds, by quantum chemistry geometry optimizations at the RI-DFT-D (RI-B97-D/TZVPP) level (Table 1; see ESI† for details, namely Fig. S1 and Table S1). The resulting optimizations have given rise to 25 configurations of comparable energetics (within less than 3.3 kcal mol⁻¹). Most of these conformations are compact forms, stabilized by interactions binding the two ends of the molecule, either a C₁₄ H-bond or an NH- π H-bond between the NH of the first Ala residue and the phenyl ring.

The 3 and 4 H-bond conformations obtained have been sorted into six families according to the topological properties of their folding (Fig. S1 and Table S1†).

- A first set, labeled **1**, is composed of structures exhibiting a 14-7L-x-10 or a 14-7L-7D-x network (Fig. 1c). They share the same C₁₄ H-bond and one C_{7L} interaction on the N terminal (γ_L turn, with the Me side chain group in an *equatorial* position relative to the C₇ ring). The diversity in the family arises from the C terminal region: the third H-bond can be either a C₁₀ bond (type II' β -turn) or a C_{7Dx} interaction (γ_D turn, with the Me group in an *axial* position relative to the C₇ ring). These topologically close conformations are labelled **1 β** and **1 γ** , respectively, according to the nature of this third bond. Besides these forms, closely related 4 H-bond stable conformations have also been found in the optimizations. They are variant structures of mixed character between the two previous ones, with the simultaneous occurrence of consecutive bifurcated and elongated C₇/C₁₀ H-bonds, giving

rise to a 14-7L-[7D-10] pattern, noted **1m**; **m** standing for *mixed character* (see Fig. S1†).

- A second set, **2**, is based on 14-7D-x-10 and 14-x-10-10 networks, with one C₁₄ bond and one C₁₀ bond (type I' β -turn) on the C terminal. Again, some variability is observed: the third bond on the N terminal side is either a C_{7D} interaction or a C₁₀ (type II β -turn) interaction. Like in the previous case, mixed character structures, with elongated H-bonds were also found (14-[7D-10]-10 patterns). Finally, for these first two families **1** & **2**, a few conformations with opposite Ramachandran (ϕ , ψ) angles (backbone mirror symmetry) are also found to be stable and are indicated by prime symbols 1' and 2'.

- Family **3** is composed of triple C₇ (x-7-7-7) structures.

- The remaining structures exhibit 4 H-bonds and are sorted into two distinct families: Set **4** corresponds to triple C₇ bonds completed with either a C₁₄ or a π H-bond linking the two ends of the molecule. Family **5** exhibits 4 interactions and its backbone is organized around a C₁₁ interaction (see Fig. S1†). A few very stable conformations (**4A**, **5B**) are found among these families. Finally a **6th** type of backbone (3₁₀ C₁₀-C₁₀ helix) has also been added in Tables 1 and S1† for the sake of discussion.

Since NH stretch vibrational spectroscopy is very sensitive to the molecular H-bonding, comparison of experiment with synthetic IR spectra enables us to propose a structural assignment. Theoretical spectra have been obtained from scaled harmonic frequencies obtained at the RI-B97-D level of theory. This method was recently validated (with a TZVP basis) on a series of NH stretch data on short peptides.²¹ With the presently used TZVPP basis set, a scaling factor of 0.9744, adjusted over 28 benchmark NH stretches in 19 different peptide conformations⁶⁻¹⁸ leads to a standard deviation (SD, root mean square deviation) of 7 cm⁻¹ and a maximum deviation of 17 cm⁻¹ over the whole series.

With the present molecule, the analysis of the standard deviations of the 3 most red-shifted calculated frequencies from the 3 experimentally measured for A and B (Table 1) shows unambiguously that 3-H-bond conformations from sets **2** and **3** lead to SD larger than 22 cm⁻¹ and maximum deviation larger than 24 cm⁻¹. For this reason they can be ruled out. As expected for species having four H-bonds, disagreement is even worse (**4-5** sets).

The assignment of A and B has therefore to be found within set **1**. The mixed structures of this family, which exhibit bifurcated H-bonds (**1m1-b**, **1m2-b**, **1'm-a**), have to be ruled out because their elongated C₇/C₁₀ H-bond pair causes too small red shifts, in spite a few of these conformations are quite stable (for instance **1m1-b**, which is the absolute minimum on the RI-B97-D surface). According to the SD scores in Table 1, the 3 most compatible candidates (**1 β -a**, **-b** and **-c**) are those conformations (Fig. 2) which resemble the extremity of a β -hairpin,²² with antiparallel C₁₀ and C₁₄ H-bonds and are further stabilized by a C_{7L} H-bond on the N terminal. The next candidate, **1 γ -a**, was discarded owing to its significantly lesser quality fit, as testified by its SD and maximum deviation, compared with those reached on the benchmark molecules.

The three **1 β** conformations differ mainly by the arrangement of the benzyl tail; the most stable correspond to dispersion stabilized contacts between phenyl and methyl groups, which

Table 1 Conformations of Ac-(Ala)₄-O-Bzl, as obtained after potential surface exploration (see ESI†) and subsequent energy optimization at the RI-B97-D/TZVPP level of theory, sorted according to their H-bond network. 0 K enthalpies and 300 K free energy are given in kcal mol⁻¹. The rms deviations (SD) between exp. and the 3 most red IR NH stretch calculated frequencies are given in cm⁻¹, for both A and B conformers, together with the maximum discrepancy for these 3 frequencies

Conformation type		ΔH 0 K kcal mol ⁻¹	ΔG 300 K kcal mol ⁻¹	A		B	
				$\Delta\nu$ rms cm ⁻¹	$\Delta\nu$ max cm ⁻¹	$\Delta\nu$ rms cm ⁻¹	$\Delta\nu$ max cm ⁻¹
14-7L-x-10(II')	1β-a	0.23	0.00	16	25	13	21
	1β-b	1.25	1.73	7	9	14	21
	1β-c	2.53	1.30	16	20	18	23
14-7L-[7D-10(II')]	1m1-b	0.00	0.62	23	36	37	56
	1m2-b	1.97	3.75	46	78	59	98
14-7L-7D-x	1γ-a	1.72	1.88	19	28	31	49
14-7D-[7L-10(II)]	1'm-a	1.98	4.42	63	77	63	80
14-[7L-10(II)]-10(I')	2m-a	0.15	1.18	44	65	30	44
	2m-b	0.92	1.39	44	63	30	42
14-7L-x-10(I')	2β-a	2.67	2.87	73	102	60	82
14-7D-x-10(I)	2'γ-a	2.07	2.40	27	40	41	61
14-X-10(II')-10(I)	2'β-a	3.01	4.77	37	62	25	42
14-[7D-10(II')]-10(I)	2'm-a	0.95	1.80	40	54	25	33
	2'm2-b	1.74	3.23	82	102	66	82
	2'm3-c	2.05	2.20	20	27	19	27
x-7L-7L-7L	3A	2.57	2.25	31	51	34	51
x-7L-7D-7L	3B	3.30	3.70	31	52	44	73
π -7D-7L-7D	4A	0.06	1.95	101	129	113	129
14-7D-7L-7D	4B	1.77	2.45	45	52	58	67
π -7L-7D-7L	4'A-a	1.21	2.92	64	94	73	94
π -7L-7D-7L	4'A2-b	1.51	3.30	75	85	88	100
14-7L-7D-7L	4'B	1.65	2.80	44	57	54	71
π -11-10(I)-7D	5A	0.07	0.83	35	56	43	56
5-11-7D-7L	5B	1.85	1.51	42	63	38	63
x-x-10(I)-10(I)	6	2.90	1.27	98	98	84	113

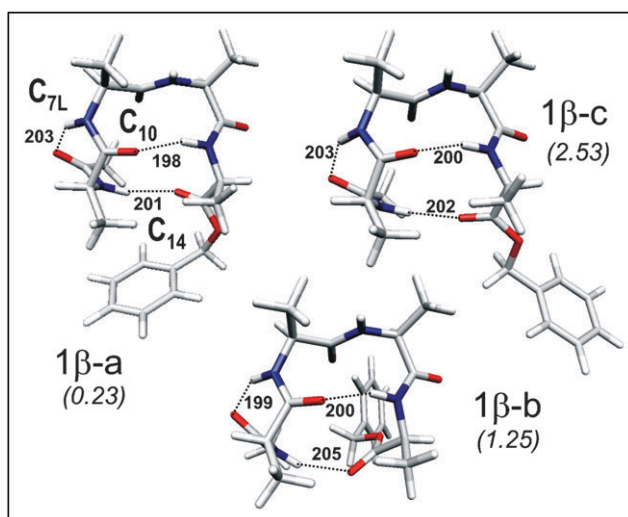


Fig. 2 RI-B97-D/TZVPP structures of the three relevant conformations of the Ac-(Ala)₄-O-Bzl molecule (Family 1) for the assignment of the IR/UV spectra. H-bond distances are given in pm. The 0 K enthalpies (in kcal mol⁻¹ in brackets) are given relative to the lowest energy conformer (**1m1-b**, see Table 1).

are not accounted for by a standard functional like B3LYP.²³ The slight variations of the backbone geometry with the tail orientation (Fig. 2) illustrate the robustness of the structure with its environment. These variations are however significant

enough to be revealed by NH stretch vibrational spectroscopy. Comparison with experiment (see SD in Table 1 and Fig. 1c) suggests that **1 β -b** is by far the best candidate for A. For B, competition is more open, and assuming that **1 β -b** fits A, B should be assigned to either **1 β -a** or **1 β -c**.

This assignment is consistent with the apparently different chromophore environments between A and B as suggested by the significant UV shift between the origin bands of the two conformers (Fig. 1a).

One should notice the challenging character of the **1 β** conformations, where the backbone structure results from a subtle balance between strong H-bonds competing with each other and possibly influenced by dispersion forces. In this context, finding structures capable to reproduce the experiment (with small SD: < 14 cm⁻¹) illustrates the potentialities of the method employed, at a relatively modest calculation cost, which makes it affordable for larger species.

A careful analysis of the data however also demonstrates its limits. Discrepancies as large as ~20 cm⁻¹ can still be observed: we assign them to an imperfect account for the competition between H-bonds mentioned above. In the same line, one can remark that fine details of the vibrational structure (Fig. 1b) are not properly reproduced, for instance the 20 cm⁻¹ spectral shift in the central C₁₀ IR band between conformers A and B.

Another point concerns the mixed character structures, with bifurcated C₁₀/C₇ H-bonds, which are found to be rather

stable (**1m**, **1'm**, **2m**, **2'm** structures) but nevertheless not observed experimentally. Limited optimization tests have been carried out at the RI-MP2/SVP level. It has been shown that **1m1b** is no longer a minimum and converges to another **1β** form, which, in its turn, is not a RI-B97-D minimum. In spite of the drawbacks of the MP2 method performed with a so small basis set, this casts some doubts about the reality of these bifurcated structures as real minima of the system. It also suggests that the position of the benzyl tail may not be properly described in detail at the RI-B97-D level.

Finally, we have to raise the question of the accuracy of the calculated energetics. One has to notice that, apart from the **1β** family, several rather stable minima with structures not compatible with the observation are also found on the surface (**1m1-b**, **2m-a**, **2'm-a**, **4A**, **5A**). Owing to their 0 K stability, one would expect to observe them, which suggests that the RI-B97-D relative error on energetics might be on the order of 2 kcal mol⁻¹ in such a system. One has, however, to consider the specific experimental conditions involving both laser desorption coupled with a supersonic expansion, in which the final population might be more properly described from high temperature free energies rather than 0 K enthalpies, because of the fast tempering caused by the expansion as suggested by a few reports.^{24,25} In this respect, Table 1 shows that most of these stable forms become much less favored at room temperature leading to a complete conformational reordering. This thermal effect is mainly due to the density of accessible vibrational modes at a finite temperature, which disfavors the rigid species to the benefit of the loose conformations, which have a greater density of low frequency modes.

In this respect the entropy effect disfavors locked systems (for instance: 4 H-bond conformations or **1β-b**, with its tail embedded in the backbone fold) to the benefit of more floppy systems (the loose **3₁₀** helix, **6**, with only two H-bonds; **1β-c** with its loose tail).

A combination of these two effects (energetics inaccuracy and entropy effect) could be responsible for the absence in the expansion of stable forms according to 0 K calculations.

Obviously these β-hairpin model peptide systems, with a large content of dispersive interactions and a folding pattern characterized by a tight network of competitive H-bonds, turn out to be very stringent tests for the new and efficient quantum chemistry methods currently being developed. The qualitative energetics obtained with the presently used RI-B97-D method together with the quantitative agreement on vibrational spectroscopy nevertheless suggests it should greatly help experimentalists to tackle the folding of larger peptide species under isolated conditions.

Acknowledgements

The present work is supported by the french *Agence Nationale de la Recherche* under the contract ANR-08-BLAN-0158-01.

References

- 1 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
- 2 H. Valdes, V. Spiwok, J. Rezac, D. Reha, A. G. Abo-Riziq, M. S. de Vries and P. Hobza, *Chem.–Eur. J.*, 2008, **14**, 4886–4898.
- 3 T. Haber, K. Seefeld, G. Engler, S. Grimme and K. Kleinermanns, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2844–2851.
- 4 T. van Mourik, *J. Chem. Theory Comput.*, 2008, **4**, 1610–1619.
- 5 Special issue on “Stacking interactions”, *Phys. Chem. Chem. Phys.*, 2008, **10**, Issue 19, in particular: H. Valdes, K. Pluhackova, M. Pitonak, J. Rezac and P. Hobza, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2747–2757.
- 6 I. Compagnon, J. Oomens, J. Bakker, G. Meijer and G. von Helden, *Phys. Chem. Chem. Phys.*, 2005, **7**, 13–15.
- 7 J. M. Bakker, C. Plutzer, I. Hunig, T. Haber, I. Compagnon, G. von Helden, G. Meijer and K. Kleinermanns, *ChemPhysChem*, 2005, **6**, 120–128.
- 8 W. Chin, J. P. Dognon, C. Canuel, F. Piuze, I. Dimicoli, M. Mons, I. Compagnon, G. von Helden and G. Meijer, *J. Chem. Phys.*, 2005, **122**, 054317.
- 9 W. Chin, I. Compagnon, J. P. Dognon, C. Canuel, F. Piuze, I. Dimicoli, G. von Helden, G. Meijer and M. Mons, *J. Am. Chem. Soc.*, 2005, **127**, 1388–1389.
- 10 W. Chin, F. Piuze, I. Dimicoli and M. Mons, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1033–1048.
- 11 I. Compagnon, J. Oomens, G. Meijer and G. von Helden, *J. Am. Chem. Soc.*, 2006, **128**, 3592–3597.
- 12 H. Fricke, G. Schafer, T. Schrader and M. Gerhards, *Phys. Chem. Chem. Phys.*, 2007, **9**, 4592–4597.
- 13 H. Fricke, A. Funk, T. Schrader and M. Gerhards, *J. Am. Chem. Soc.*, 2008, **130**, 4692–4698.
- 14 R. A. Jockusch, F. O. Talbot, P. S. Rogers, M. I. Simone, G. W. J. Fleet and J. P. Simons, *J. Am. Chem. Soc.*, 2006, **128**, 16771–16777.
- 15 E. E. Baquero, W. H. James, S. H. Choi, S. H. Gellman and T. S. Zwier, *J. Am. Chem. Soc.*, 2008, **130**, 4784–4794.
- 16 W. Chin, F. Piuze, J. P. Dognon, L. Dimicoli, B. Tardivel and M. Mons, *J. Am. Chem. Soc.*, 2005, **127**, 11900–11901.
- 17 V. Brenner, F. Piuze, I. Dimicoli, B. Tardivel and M. Mons, *J. Phys. Chem. A*, 2007, **111**, 7347–7354.
- 18 J. A. Stearns, O. V. Boyarkin and T. R. Rizzo, *J. Am. Chem. Soc.*, 2007, **129**, 13820–13821.
- 19 T. D. Vaden, S. A. N. Gowers, T. de Boer, J. D. Steill, J. Oomens and L. C. Snoek, *J. Am. Chem. Soc.*, 2008, **130**, 14640–14650.
- 20 F. Piuze, I. Dimicoli, M. Mons, B. Tardivel and Q. Zhao, *Chem. Phys. Lett.*, 2000, **320**, 282–288.
- 21 Y. Bouteiller, J. C. Pouilly, C. Desfrancois and G. Grégoire, *J. Phys. Chem. A*, 2009, **113**, 6301–6307.
- 22 A. J. Maynard, G. J. Sharman and M. S. Searle, *J. Am. Chem. Soc.*, 1998, **120**, 1996–2007.
- 23 The relative electronic energies of the **1β a-c** forms optimized at the B3LYP/6-31+G* level are 0, 0.74 and –0.20 kcal mol⁻¹, respectively (taking **1β-a** as a reference), to be compared to 0, 0.93 and 2.52 kcal mol⁻¹ at the B97-D/TZVPP level.
- 24 R. S. Ruoff, T. D. Klots, T. Emilsson and H. S. Gutowsky, *J. Chem. Phys.*, 1990, **93**, 3142–3150.
- 25 D. Reha, H. Valdes, J. Vondrasek, P. Hobza, A. Abu-Riziq, B. Crews and M. S. de Vries, *Chem.–Eur. J.*, 2005, **11**, 6803–6817.