# *Ab-Initio* calculations of proline vibrations with and without water, consequences on the infrared spectra of proline-rich proteins

A. Banc*, B. Desbat†, D. Cavagnat‡

* Centre de Recherche Paul-Pascal, Université Bordeaux-1 CNRS, UPR 8641, 115 avenue A. Schweitzer, F-33600 Pessac, France.
Current adress: Université Montpellier 2, Laboratoire Charles Coulomb UMR 5221, F-34095, Montpellier, France. CNRS, Laboratoire Charles Coulomb UMR 5221, F-34095, Montpellier, France.

† Laboratoire de Chimie et Biologie des Membranes et Nanoobjets, Université Bordeaux-1, UMR 5248-CNRS, ENITAB, 2 rue Robert Escarpit, F-33607 Pessac, France.

‡ Institut des Sciences Moléculaires, Université Bordeaux-1, CNRS, UMR 5255, F-33405 Talence, France.

**Abstract:** The infrared spectra of proline rich proteins display a strong band in the 1450 $cm^{-1}$ region. In the literature, this band was assigned either to the deformation modes of the $CH_2$ and $CH_3$ groups or to the CN stretching mode of proline residues. In order to establish the correct assignment of this band, the impact of proline vibrations in a polypeptide chain is studied and *ab-initio* calculations are performed for a model molecule (I) containing a repeat unit of polyproline. A strong band is effectively calculated in the 1450 $cm^{-1}$ region and mostly assigned to CN stretching whereas, due to the absence of N-H bond, there is no amide II band. These results are in good agreement with the spectral features observed in the FTIR spectra of gliadins. Moreover, the spectral shifts calculated when a water molecule is complexed with (I) are consistent with the hydration effect observed in the experimental data.

**Key words:** ab-initio calculations - prolines - vibration mode- gliadins.

**Introduction:** Proline residues play an important role in the properties and in the structure of globular and fibrous proteins. The absence of amino hydrogen in the chain backbone prevents H-bonded interactions, while the presence of five-membered pyrollidine ring constrains the rotation angle of the $C^{\alpha}$-N bond. Hence, this residue is often found at bends in the polypeptide chain. Moreover, polyproline can adopt two helical conformations: the form I (PPI) characterized by a right handed helix, in which each peptide bond is in a *cis* configuration, and the form II (PPII), a left handed helix where the peptide bonds are in the *trans* configuration[1]. In collagen, prolines seem to pre-organize the individual strands in a PPII conformation to form the typical collagen triple helix[2]. In transmembrane $\alpha$ helices, proline motifs are believed to generate conformational switches at the origin of the translocation of solutes[3]. It has also been suggested that proline residues may play an important role in the folding of membrane proteins[4]. Prolines are also important in prolamins, a class of storage protein of cereal seeds, characterized by their high content of prolines and glutamines (up to 30 mol % and 50 mol % respectively)[5]. These residues are mostly found in repetitive sequences which describe the repetitive domain of these proteins. Spectroscopic studies[6, 7, 8, 9, 10, 11] indicate that, contrary to the globular non-repetitive domains characterized by classical secondary structures ($\alpha$-helices, $\beta$-sheet), repetitive domains are low folded and subjected to secondary structure changes with the environment (solvent, temperature, pressure…). In infrared studies[12, 13], authors focused on amide I band and assigned the strong band in the 1450 $cm^{-1}$ region to CH deformation. However, in other studies on bacteriorhodopsin[14, 15] and polyproline[16], authors assigned this strong band to the C-N vibrations of proline. In the 1450 $cm^{-1}$ region, polyproline infrared spectra exhibit a strong band varying in frequency and intensity according to hydration level[17] and temperature[18]. De Lozé and Josien[17] suggested that these variations were due to formation, by hydrogen bonding, of complex between polyproline and water. Moreover, Caswell and Spiro[16] showed by ultraviolet resonance

Raman spectroscopy that polyproline I and polyproline II structures exhibit two different vibration modes (1435 $cm^{-1}$ and 1465 $cm^{-1}$ respectively), that they assigned to cis and trans conformations of imide bonds contained in each structure.

In order to establish the 1450 $cm^{-1}$ band assignment in gliadin infrared spectra, we chose to perform *ab initio* calculations for a model molecule, $C_7H_{13}NO$, containing a repeat unit of polyproline (noted **I**) (Figure 1). This molecule is isomorphous with the structure of cis L proline residue in natural proteins. This simplification is acceptable as we are interested by the vibrations of the proline cycle, which cannot be strongly perturbed by the exchange of the $CH_2$ group extremities for $CH_3$ groups. The calculations were firstly done for isolated molecule **I.** Similar calculations were then done for a complex (I)-$H_2O$ in order to analyse the hydratation effect observed in previous experimental results obtained for polyproline.

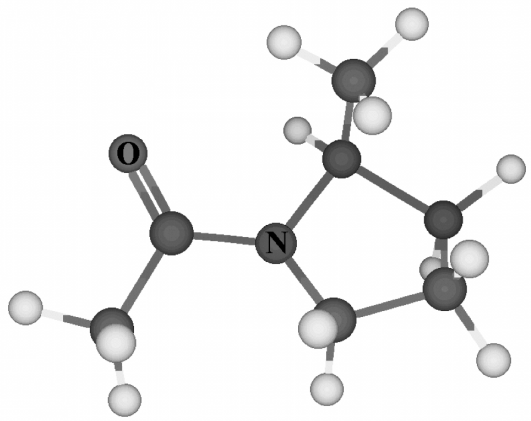


Figure 1. Molecule **I**

**Calculations Method:** The geometry optimizations, vibrational frequencies and absorption intensities, were calculated by Gaussian 03 program[22] on a SGI Altix XE 1300 (with four processors) of the Pôle Modélisation of the Institut des Sciences Moléculaires (University Bordeaux I). Calculations of the optimized geometries were performed at the density functional theory level using B3PW91 functional and cc-pVTZ basis set. Vibrational

frequencies and IR intensities were calculated at the same level of theory. For comparison to experiment, the calculated frequencies were scaled by 0.98.

**Results:** First of all, we display the infrared spectra of gamma and omega gliadins, two wheat prolamins, in the amide frequency domain (Fig.2). The two proteins present the characteristic amide I and II absorptions, but also one band around 1450 $cm^{-1}$ near the $CH_2$ and $CH_3$ absorptions. This band is relatively stronger than the bands observed for the other proteins in the same range of frequencies. Moreover, the omega gliadin which contains a higher proportion of prolines displays a lower relative intensity for the amide II and a stronger intensity for the band at 1450 $cm^{-1}$.

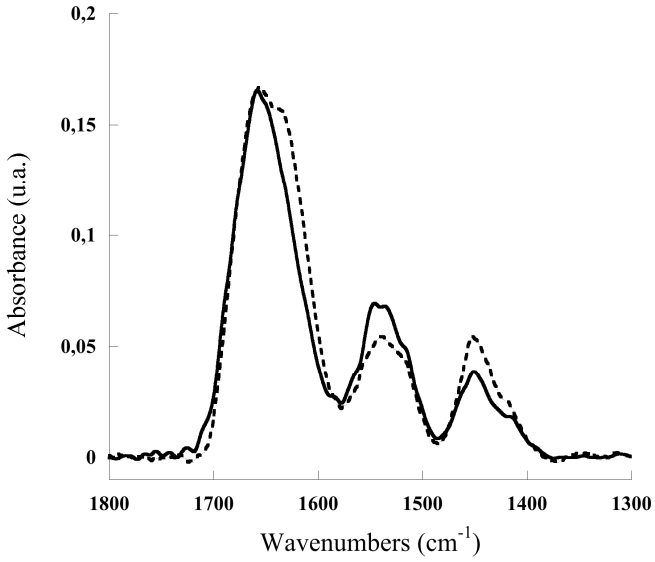


Figure 2. Infrared spectra of lyophilized gamma (continuous line) and omega (dash line) gliadins[11]

The evolution of the 1450 $cm^{-1}$ and of the amide II band peak areas is directly correlated with the proline content in both protein sequences[11] (17.5% of proline in γ-gliadin[19] and 29% in ω-gliadin[20]).

In Table 1, we report the frequencies, infrared intensities and assignment of the B3PW91/cc-pVTZ calculated bands of the isolated molecule **I** in the amide frequency domain**.** Figure 3a displays the optimized geometry of this molecule.

Only two strong bands are calculated in this domain. The first one is essentially due to the C=O stretching vibration and is assigned to the amide I band. The second one involves mostly the C-N stretching vibration. In agreement with the structure of the proline residue in the peptidic chain, no amide II band (coupled ν(C-N), δ(N-H) vibrations) is expected, as there is no N-H bond on the tertiary amide group. As a consequence, the decrease of the amide II intensity and the presence of a strong band at lower frequency (ν C-N) in proline rich protein FTIR spectra is directly related with the lower proportion of NH groups in these proteins. This effect can be observed for gliadins but also for collagen proteins, which contain also a high proportion of proline residues.

| Frequency | Infrared intensity $I^b$ | Assignment |
|---|---|---|
| 1739.5<br>1704.7* | 322.6 | ν $C_2$=$O_3$ (79 %) ν $C_2$-$N_1$ (7 %) |
| 1436.3<br>1407.6* | 342.5 | ν $C_2$-$N_1$ (45 %) ν $C_2$-$C_4$ (14 %) |
| 1520.2<br>1489.8* | 1.3 | δ$C_8H_{17}H_{16}$ (85.7%) |
| 1494.1<br>1464.3* | 3.6 | δ$C_9H_{18}H_{19}$ (49.5 %) δ$C_{10}H_{20}H_{21}$ (30.8 %) |
| 1484.6<br>1454.9* | 2.6 | δ$C_9H_{18}H_{19}$ (29.2 %) δ$C_{10}H_{20}H_{21}$ (43 %) |

Table1. B3PW91/cc-pVTZ calculated frequencies ($cm^{-1}$) **(*** scaled by 0.98)
and infrared intensities (KM/Mole) of the isolated molecule **I** in the 1800-1400 $cm^{-1}$ domain.

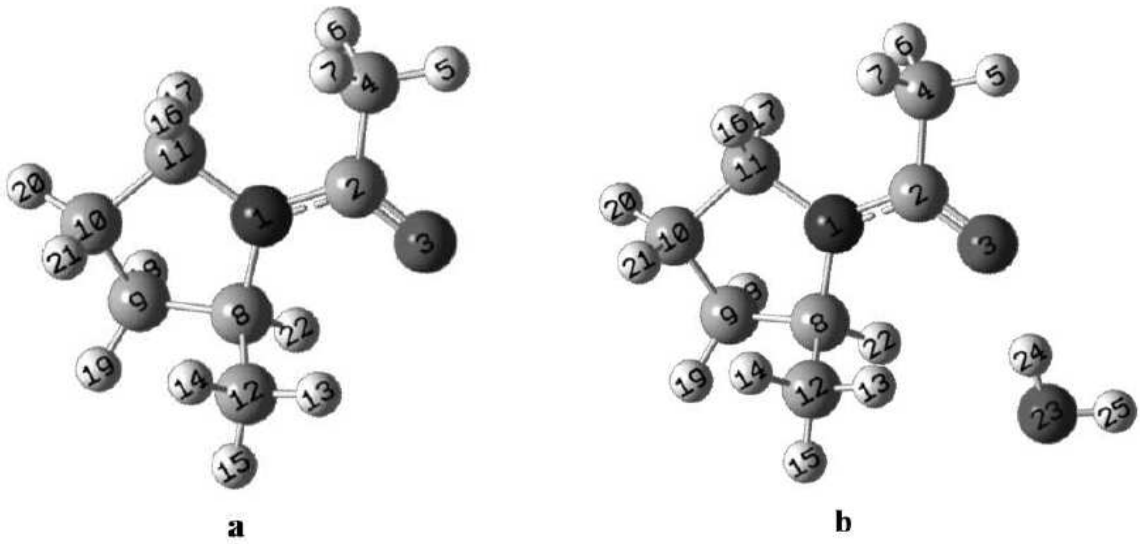


Figure 3. B3PW91/cc-pVTZ optimized geometries of isolated molecule (I) (a) and of the molecule (I)- water complex (b)

| Isolated molecule (I) | | Molecule (I) + $H_2O$ | | $\Delta \nu$ |
|---|---|---|---|---|
| Frequency | Intensity I | Frequency | Intensity I | |
| 1739.5<br>1704.7* | 322.6 | 1713.0<br>1678.7* | 430.0 | -26.5<br>-26.0* |
| 1436.3<br>1407.6* | 342.5 | 1451.9<br>1422.9*<br>1466.4<br>1437.1* | 223.8<br>102.8 | +15.6<br>+15.3*<br>+30.1<br>+29.5* |

Table 2. Comparison between calculated frequencies ($cm^{-1}$) (***** scaled by 0.98) and infrared intensities (KM/Mole) of the two strongest bands of molecule (I) without and in presence of a water molecule

The B3PW91/cc-pVTZ optimized geometries of molecule (I)-$H_2O$ complex (figure 3b) shows that the water molecule is hydrogen bonded to the oxygen of the carbonyl. This produces an important modification of the amide frequencies. Indeed, as shown in Table 2,
the amide I displays a negative shift of -26 $cm^{-1}$ and its intensity increases notably in comparison with the vibration of the isolated molecule. Moreover, the ν(C-N) band splits in two close bands with almost the same intensity, but their frequencies are shifted to higher frequencies (+15 and +30 $cm^{-1}$). These results are in good agreement with the experimental spectra obtained for poly-L-proline. With hydration, Wellner et al[12] measured a shift of -26 $cm^{-1}$ for amide I and +29 $cm^{-1}$ for the CN vibration mode, and Johnston et al[21] reported shifts of -24 $cm^{-1}$ and +37 $cm^{-1}$ respectively. Moreover, when increasing temperature, Swenson et al[18] observed an evolution of frequencies consistent with a decrease of the hydrogen bond strength.

The analysis of the calculated atomic Mulliken charges of molecule (I) reported in table 3 shows that oxygen (**3**) brings the higher negative charge whereas nitrogen (**1**) brings the higher positive charge. Thus, the hydrogen bond with water molecule is preferentially done with the oxygen of the carbonyl group. Moreover, the effect of the hydrogen bonded water increases significantly the charges separation between the different atoms and confirms the possibility to get a modification of the amide group vibrational frequencies by the formation of a hydrated system.

| Molecule(**I**) | | Molecule (**I**)+$H_2O$ | |
|---|---|---|---|
| atoms | Mulliken atomic charges | atoms | Mulliken atomic charges |
| N1 | .542187 | N1 | .549348 |
| C2 | .078766 | C2 | .049060 |
| O3 | -.847570 | O3 | -.924000 |
| C4 | -.455120 | C4 | -.439108 |
| H5 | .159279 | H5 | .202797 |
| H6 | .214123 | H6 | .195968 |
| H7 | .217068 | H7 | .201374 |
| C8 | .242173 | C8 | .410996 |
| C9 | -.392007 | C9 | -.460822 |
| C10 | -.407759 | C10 | -.388896 |
| C11 | -.309836 | C11 | -.351182 |
| C12 | -.857828 | C12 | -.900719 |

Table 3. Calculated Mulliken charges for Molecule (**I)** and complex (**I)** + H2O

**Conclusion:** In summary, the vibration mode observed around 1450 $cm^{-1}$ for proline rich proteins is characteristic of the coupled ν(C-N), ν(C-C) vibrations of the tertiary amide. The position of this band is sensitive to the hydration state of the oxygen amide group and shifts to higher frequencies when the hydrogen bonds are formed. Moreover, a high proportion of proline residues in one protein produces also a significant decrease of the relative intensity of the amide II band in comparison with the amide I band.

References:


1 Mandelkern L., *in Poly-alpha-aminoacids*, chapter 13, G.D. Fasman, editor, Marcel Dekker, New-York (1967)

2 M. D. Shoulders and R. T. Raines, Annual Review of Biochemistry **78**, 929–958 (2009).

3 F. Cordes, J. Bright, and M. Sansom, Journal of Molecular Biology **323**, 951–960 (2002).

4 H. Lu, T. Marti, and P. Booth, Journal of Molecular Biology **308**, 437–446 (2001).

5 P. Shewry and A. Tatham, Biochemical Journal **267**, 1–12 (1990).

6 E. Blanch, D. Kasarda, L. Hecht, K. Nielsen, and L. Barron, Biochemistry **42**, 5665–5673 (2003).

7 M. Pezolet, S. Bonenfant, F. Dousseau, and Y. Popineau, FEBS Letters **299**, 247–250 (1992).

8 A. Tatham, P. Masson, and Y. Popineau, Journal of Cereal Science **11**, 1–13 (1990).

9 D. Georget and P. Belton, Biomacromolecules **7**, 469–475 (2006).

10 N. Wellner, E. Mills, G. Brownsey, R. Wilson, N. Brown, J. Freeman, N. Halford, P. Shewry, and P. Belton, Biomacromolecules **6**, 255–261 (2005).

11 A. Banc, B. Desbat, D. Renard, Y. Popineau, U. Mangavel, and L. Navailles, Langmuir **23**, 13066–13075 (2007).

12 N. Wellner, P. Belton, and A. Tatham, Biochemical Journal **319**, 741–747 (1996).

13 F. Secundo and N. Guerrieri, Journal of Agricultural and Food Chemistry **53**, 1757–1764 (2005).

14 K. Gerwert, B. Hess, and M. Engelhard, FEBS Letters **261**, 449–454 (1990).

15 K. Rothschild, Y. He, D. Gray, P. Roepe, S. Pelletier, R. Brown, and J. Herzfeld, Proceedings of the National Academy of Sciences of the United States of America **86**, 9832–9835 (1989).

16 D. Caswell and T. Spiro, Journal of the American Society **109**, 2796–2800 (1987).

17 C. Deloze and M. Josien, Biopolymers **8**, 449– (1969).

18 C. Swenson and R. Formanek, Journal of Physical Chemistry **71**, 4073– (1967).

19 Y. Popineau and F. Pineau, Journal of Cereal Science **3**, 363–378 (1985).

20 Y. Popineau, J. Leguerroue, and F. Pineau, Lebensmittel-Wissenschaft and Technologie **19**, 266–271 (1986)

21 N. Johnston and S. Krimm, Biopolymers **10**, 2597–2605 (1971).

22 Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Jr .Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Naktasuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J.E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Statmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, D. J.; Fox, T.; Keith, M. A.; Al-Laham, C. Y.; Peng, A.; Nanayakkara, M.; Challacombe, R. L.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.04, Gaussian Inc., Pittsburgh, PA, **2003**.