ELECTRONIC FACTORS FAVOURING THE CIS CONFORMATION IN PROLINE PEPTIDIC BONDS

Rivail, J. L. ; Bouchy, A.; Loos, P. F.

Equipe de Chimie et Biochimie Théoriques, UMR 7565 CNRS-Université Henri Poincaré
B.P.239, 54506 Nancy-Vandoeuvre, France.

Fax : +33 383 68 43 71, E-Mail:Jean-Louis.Rivail@cbt.uhp-nancy.fr

Abstract

A quantum chemical study based on B3LYP, 6-31+G* computations offered the possibility of comparing three dipeptides containing as a central residue proline, N ethyl alanine and an alanine respectively in their cis and trans conformations. The presence of the electron donating alkyl substituent on the nitrogen atom is an important factor favouring the cis conformation. The stabilisation of the cis conformer is interpreted as being due to electrostatic and induction interactions between the carbonyl group and the substituent on the nitrogen atom.

Resumen

Un estudio químico cuántico basado en cálculos computacionales B3LYP, 6.31+G* permitió la comparación de tres dipéptidos que contienen como residuo central prolina, N etil alanina y alanina, respectivamente, en sus conformaciones cis y trans. La presencia de un sustituyente alquilo donador de electrones sobre el átomo de nitrógeno es un factor importante que favorece la conformación cis. La estabilización del confórmero cis es debida a interacciones electrostáticas y de inducción entre el grupo carbonilo y el sustituyente sobre el átomo de nitrógeno.

Introduction

Protein folding is one of the major problems that chemistry and biochemistry have to face at the beginning of the 21st century. Among the information to be gained to rationalize the phenomenon, peptidic conformation plays a central role, and I.G. Csizmadia and coworkers made an important contribution to the understanding of what may influence the three dimensional structure of a protein at the level of the peptidic bonds [1]. Although the great majority of the peptidic bonds in proteins adopt the trans conformation, it happens that a few of them take the other possible structure: the cis conformation [2]. This occurrence is particularly frequent when the bond involves a proline residue bonded at its amino moiety [3]. This remark suggests that, in addition to the many factors which may be responsible for these unusual structures, electrostatic or steric interactions with the surrounding parts of the protein, proline may have some intrinsic properties which favour such conformations. Some theoretical investigations have already been

done on the conformations of the proline peptidic bond [4-6], which confirm the observed tendency, but to our knowledge, no convincing explanation has been proposed to rationalise this result. The aim of this study is essentially a qualitative investigation of the electronic effects which may help understanding the special role of proline in proteins. Several effects may be invoked to explain this particular behaviour compared to the other aminoacids. One immediately thinks of a possible modification of the electronic properties of the nitrogen atom induced by the five membered ring which constrains the bond angles and also of a possible modification of the local electronic distribution due to the electron donating properties of the (cyclo) alkyl group bonded to the nitrogen atom. Interactions between non bonded atoms may also occur. The flexibility of the five membered ring does not seem to be a major factor since there are several examples of cis-proline bonds in which the conformation of the ring is changed. This is the case, for instance, of the cis peptidic bonds between Tyr 92 and Pro 93 or Asn 113 and Pro 114 in ribonuclease A [7].
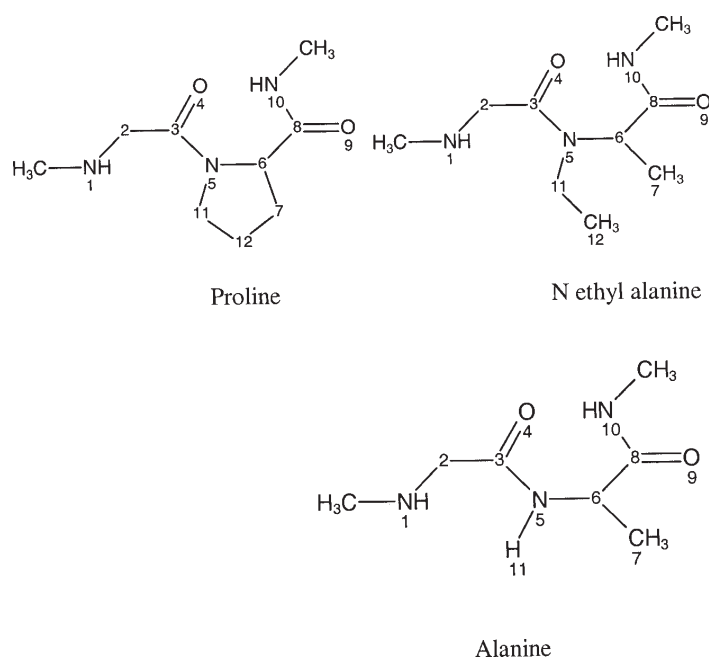
Scheme 1: Structure of the dipeptides (trans conformers)

In order to gain some insight in the cis-proline conformation, we performed a series of computations on a model molecule with a central glycine-proline peptidic bond and compared the results with those obtained when proline is replaced by alanine and by the hypothetical N ethyl alanine which mimics the influence of the alkyl chain free of ring constrains. The reference model system is the N methyl glycine–proline methylamide (scheme 1).

Computations

The computations have been performed with the Gaussian 03 set of programs [8]. Owing to the aim of this study, an intermediate computational level has been adopted: DFT formalism with the B3LYP exchange-correlation functional [9] and the 6-31+G* standard basis set. The

geometries are fully optimized and the nature of the stationary points assessed by computing the vibrational frequencies. The electronic population is analysed by means of Natural Bond Orbitals (NBO) [10] and Mayer bond orders [11].
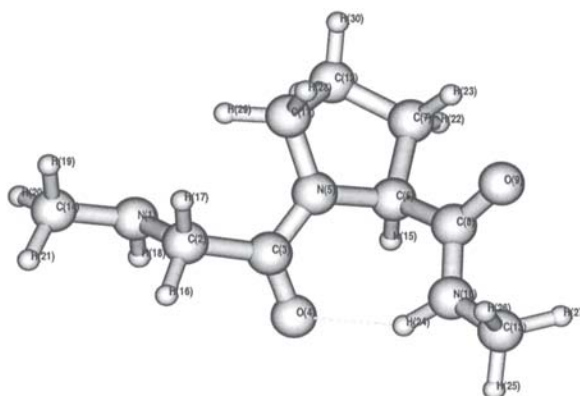
Results

Figure 1: Three dimensional structure of the cis-proline dipeptide

Conformations and energies

The global minimum of the proline dipeptide is the trans conformation of the peptidic bond, in agreement with a previous result on a dipeptide having a slightly different composition [4]. But the cis conformer, represented on figure 1, is less than 1 kcal/mol above the global minimum (figure 2). This is also true when proline is replaced by N ethyl alanine. On the contrary, the difference in favour of the trans conformer is noticeably greater in the case of alanine (more than 5 kcal/mol). The conformation of the chain is characterized by the angles ω, φ, ψ given on table 1. For both conformations, the backbones of the three peptides are rather similar. The only noticeable difference in the case of the cis proline is the orientation of the CO bond of the second carbonyl group, not belonging to the peptidic bond of interest, which is characterized by the ψ angle, negative in the case of cis proline and positive in both other cis conformers.
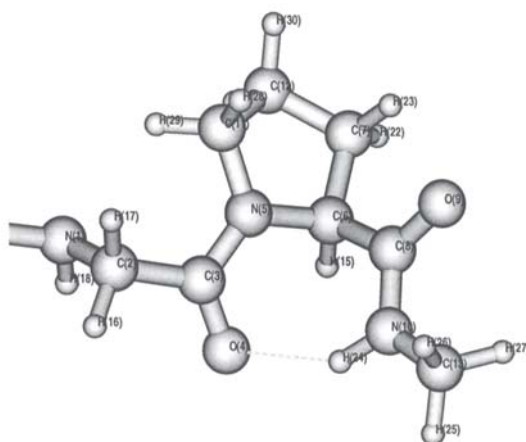

Figure 2: Three dimensional structure of the trans-proline dipeptide

This twist of *ca.* 60° may be due to constrains introduced by the five membered cycle. All the attempts to localize another minimum with a negative ψ angle failed, and similarly, the attempts to find another conformation of the other cis conformers with a positive ψ angle were unsuccessful. The φ angle indicates a flattening of the cycle in both conformations of proline: 21.52° and 8.24° for the cis and trans conformations respectively. For comparison, in tetrahydropyrole (the cycle alone), the corresponding value is 44.23°. Finally, intramolecular hydrogen bonds occur in both conformations: between the hydrogen atom bonded to nitrogen 10 and N_1 in the cis conformations and between this hydrogen atom and O_4 in the trans conformation. It is worth noticing that the latter hydrogen bond is shorter in the trans proline dipeptide than in both other trans molecules. This hydrogen bond, obviously stronger in the trans than in the cis conformation may contribute to the relative stability of the trans proline conformer. In proteins, these hydrogen bonds sometimes cannot be formed and this is probably an extra factor favouring the cis conformation.

Table 1*: Main geometrical features of the dipeptides in cis and trans conformations*

Dipeptide	Proline		N ethyl Alanine		Alanine	
	Cis	Trans	Cis	Trans	Cis	Trans
φ (7,6,5,11) (°)	-21.52	-8.24	-60.88	-47.33	-56.94	-27.56
ψ (7, 6,8,9) (°)	-33.40	11.80	26.18	29.94	21.83	13.95
ω (4,3,5,11) (°)	-3.63	169.34	1.77	166.44	2.81	-171.88
d (3,4) (Å)	1.236	1.243	1.235	1.240	1.231	1.238
d (3,5) (Å)	1.359	1.355	1.369	1.366	1.366	1.351
d (5,6) (Å)	1.476	1.486	1.477	1.485	1.468	1.467
d (N_1…H) (Å)	2.067		2.106		2.077	
d (O_4…H) (Å)		1.971		2.185		2.056
Energy (au)	-667.943914	-667.945952	-669.148263	-669.149134	-590.526331	-590.535214
Relative energy (kcal)	0.83	0.000	0.55	0.000	5.57	0.000

Table 2: *Main characteristics of the electronic structure of the proline dipeptide conformers*

		cis	trans
Mayer's Bond indices	$C_3\text{-}O_4$	1.711	1.830
	$C_3\text{-}C_2$	0.636	0.502
	$C_3\text{-}N_5$	1.200	1.299
	$N_5\text{-}C_6$	1.003	0.920
	$N_5\text{-}C_{11}$	0.794	0.774
	$C_{11}\text{-}C_{12}$	0.698	0.609
	$C_7\text{-}C_{12}$	0.638	0.624
	$C_6\text{-}C_7$	0.730	0.770
	$C_6\text{-}C_8$	0.807	0.753
Charges	C_3	0.674	0.686
	O_4	-0.649	-0.679
	N_5	-0.494	-0.476
	$(CH)_6$	0.137	0.134
	$(CH_2)_{11}$	0.245	0.238
	$(CH_2)_{12}$	0.014	0.017
	$(CH_2)_7$	0.055	0.045

Table 3: *Main characteristics of the electronic structure of the N ethyl alanine dipeptide conformers*

		cis	trans
Mayer's Bond indices	$C_3\text{-}O_4$	1.803	1.779
	$C_3\text{-}C_2$	0.712	0.442
	$C_3\text{-}N_5$	1.200	1.219
	$N_5\text{-}C_6$	0.880	0.732
	$N_5\text{-}C_{11}$	0.687	0.784
	$C_{11}\text{-}C_{12}$	0.660	0.640
	$C_6\text{-}C_7$	0.974	0.942
	$C_6\text{-}C_8$	0.805	0.822
Charges	C_3	0.672	0.698
	O_4	-0.645	-0.680
	N_5	-0.485	-0.485
	$(CH)_6$	0.123	0.137
	$(CH_3)_7$	0.055	0.056
	$(CH_2)_{11}$	0.231	0.228
	$(CH_3)_{12}$	0.029	0.026

Table 4*: Main characteristics of the electronic structure of the alanine dipeptide conformers*

		cis	trans
Mayer's Bond indices	C_3-O_4	1.800	1.832
	C_3-C_2	0.922	0.833
	C_3-N_5	1.288	1.249
	N_5-C_6	1.016	0.921
	N_5-H_{11}	0.732	0.703
	C_6-C_7	0.842	0.717
	C_6-C_8	0.784	0.805
Charges	C_3	0.667	0.673
	O_4	-0.633	-0.678
	N_5	-0.655	-0.655
	$(CH)_6$	0.102	0.114
	$(CH_3)_7$	0.061	0.059
	H_{11}	0.436	0.447

Electronic structure

The most significant data regarding the Mayer bond indices and the charges are collected in tables 2 to 4, for both conformations of the three dipeptides. In each case, the changes are not very important from one conformation to the other. The most characteristic feature is the charge on the methylene group bonded to the nitrogen atom in the case of proline and N ethyl alanine (position 11) which is strongly positive, especially in the case of cis proline (0.245), and greater than that for tetrahydropyrole in which it is only 0.163. A consequence of this intramolecular electron transfer is an increase of the negative charge on the oxygen atom.

Discussion

The above results clearly show that the properties of the backbone do not vary substantially neither from one dipeptide to another nor from the cis to the trans conformations. The presence of an electron donor substituent on the nitrogen atom favours the cis conformer, but the effect of this substituent is not sufficient to make the cis conformer more stable than the trans form. The comparison of the proline dipeptide and the N ethyl alanine one shows that the positive charge on the methylene group at position 11 is greater for proline than for N ethyl alanine in both conformations. This is probably linked to the fact that the cycle is a better electron donor than the ethyl group. One also notices that, in the case of proline, this positive charge is greater in the cis conformation than in the trans form. If now one considers the sum of the charges of the cycle: N plus $(CH)_6$ plus the three methylene groups of the cycle, one finds -0.043 for the cis conformation and -0.042 for the trans form, so that, although the charge of the methylene group at position 11

is greater in the cis conformation than in the trans position, the electron transfer does not vary noticeably from one conformation to the other. One is led to explain the main difference between these charge distributions by an induced polarization which is greater for the cis conformation due to the neighbourhood of the negative oxygen atom. This phenomenon also occurs in the conformers of the N ethyl alanine dipeptide, but there may be some additional effects due to the flexibility of the ethyl group. For instance, the distance O_4-C_{11} is 2.729 Å in the cis proline dipeptide and 2.685 Å in the parent N ethyl alanine compound. Nevertheless the polarization is larger in the first case and obviously contributes more efficiently to stabilize the cis conformer. Finally, the main factor which favours the cis conformer in the proline derivative appears to be the sum of the electrostatic and induction interactions.

Conclusion

An analysis of the structural features of the cis and trans conformations of the three dipeptides of interest clearly shows the major role of the alkyl substituent of the nitrogen atom which reduces the energy difference between both conformer to such an extent that the cis conformer becomes more stable than the trans position. This effect is obviously linked to the electron donating properties of this substituent. The consequence is that the methylene group linked to the nitrogen atom is positively charged. In the cis conformation, this group is close to the carbonyl oxygen atom so that an electrostatic interaction is a factor which stabilizes this conformation. This effect is enhanced by an additional induction interaction. Altogether these effects stabilize the cis conformation by *ca.* 5 kcal/mol with respect to the trans form, if one refers to a similar compound built with an alanine residue. This is not negligible and is obviously one factor explaining the rather frequent occurrence of cis proline conformations in proteins. But it may not be the only explanation and probably, during protein synthesis, some additional effects, such as steric constrains, may occur. The importance of a substitution on the nitrogen atom may help to design artificial amino acids for synthesising peptides with specific properties.

References

[1]   Lang, A.; Csizmadia I.G. and Perczel, A., *Proteins: Structure function and genetics*, **2005**, *58*, 571.

[2]   See for instance: Sawaya M.R.; Loop J., *Biochemistry* **1997**, *36*, 586.

[3]   Wedemeyer, W.J.; Welker E.; Scheraga H.A., *Biochemistry*, **2002**, *41*, 14637.

[4]   Sahai, M.A.; Kehoe, T.A.K.; Koo, J.C.P.; Setiadi, D.H.; Chass, G.A.; Viskolcz, B.; Penke, B.; Pai E.F.; Csizmadia, I.G., *J. Phys. Chem. A* **2005**, *109*, 2660.

[5]   Kang Y.K., *J. Phys. Chem. B* **2002**, *106*, 2074.

[6]   Kang Y.K., *J. Mol. Struct. (Theochem)* **2004**, *675*, 37.

[7]   Berisio, R.; Sica, F.; Lamzin, V.S.; Wilson, K.S.; Zagari A.; Mazzarella L., *Acta Crystallogr., Sect.D* **2002**, *58*, 441.

[8]   Gaussian 03, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; 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.; Nakatsuji, 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.; Stratmann, 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, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision B.05* **2003**, Gaussian, Inc., Pittsburgh PA.

[9]   Lee, C.; Yang W.; Parr, R.G., *Phys .Rev. B,* **1988**, *37*, 785.

       Becke, A.D., *J. Chem. Phys.* **1993**, *98*, 1372, 5648.

[10]   Carpenter J.E.; Weinhold, F., *J. Mol. Struct. (Theochem)* **1988**, *46*, 41.; Foster J. P.; Weinhold, F., *J. Am. Chem. Soc*. **1980**, *102*, 7211-8; Reed A. E.; Weinhold, F., *J. Chem. Phys*. **1983**, *78*, 4066-73; Reed A.E.; Weinhold, F., *J. Chem. Phys*. **1985**, *83*,1736-40; Reed, A.E.; Weinhold R.B.; Weinhold, F., *J. Chem. Phys* **1985**, *83*,735-46; Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev*. **1988**, *88*, 899.

[11]   Mayer, I., *Chem. Phys. Lett.* **1983**, *97*, 270. ; *Int. J. Quantum Chem*. **1986**, *29*, 73, 477.