Solid-state NMR Study Reveals Collagen I Structural Modifications of Amino Acid Side Chains upon Fibrillogenesis

Paulo de Sa Peixoto, Guillaume Laurent, Thierry Azais, Gervaise Mosser

Solid-state NMR study reveals Collagen I structural modifications of amino acid side chains upon fibrillogenesis.

Paulo De Sa Peixoto, Guillaume Laurent, Thierry Azaïs, Gervaise Mosser.

Laboratoire de Chimie de la Matière Condensée de Paris, UMR7574-UPMC/CNRS/Collège de France, UPMC, 4 place Jussieu, 75005 PARIS, France.

Running Title : fibrillogenesis of Collagen I studied by solid-state NMR

To whom correspondence should be addressed: Gervaise Mosser, Laboratoire de Chimie de la Matière Condensée de Paris, UMR7574 UPMC/CNRS/Collège de France, 75005 Paris, France. E-mail: gmosser@snv.jussieu.fr. Tel: +33 144276553, Fax: +33 144276539.

Key-Words: Collagen I; solid-state NMR; fibrillogenesis; random coiled side-chain chemical shift; imino acids

Background : Collagen fibrillogenesis is a fundamental process both *in vivo and in vitro.*
Results : Fibrillogenesis increases the heterogeneity of conformations of side chains amino acids, impacting 40% of iminoacids.
Conclusion : A temperature-induced-attractive force component comes from significant amount of the collagen amino acids.
Significance: Brings new perspectives to studies of collagen assembly and interactions with other matrix molecules.

SUMMARY

***In vivo*, collagen I, the major structural protein in human body, is found assembled into fibrils. In the present work, we study a high concentrated collagen sample in its soluble, fibrillar and denatured states using one and two dimensional {^{1}H}-^{13}C solid-state NMR spectroscopy. We interpret ^{13}C chemical shifts variations in terms of dihedral angle conformation changes. Our data show that fibrillogenesis increases the side chain and backbone structural complexity. Nevertheless, only three to five rotameric equilibria are found for each amino acid residue indicating a relatively low structural heterogeneity of collagen upon fibrillogenesis. Using side chain statistical data we calculate equilibrium constants for a great number of amino acids residues. Moreover, based on a ^{13}C quantitative spectrum we estimate the percentage of residues implicated in each equilibrium. Our data indicate that fibril formation greatly affects hydroxyproline and proline prolyl pucker ring conformation. Finally, we discuss the implication of these structural data and propose a model in which the attractive force of fibrillogenesis comes from a structural reorganization of 10% to 15% of the amino acids. These results allow us to further understand collagen's self-assembling process and fibrillar structure.**

Collagen I, the most abundant protein in human body, is a very long (300 nm) and thin (1.5 nm) protein. It is formed by three left-handed polypeptide chains super-coiled into a right-handed helix. These polypeptide chains, composed of about 1040 amino acids each (1), are formed almost exclusively by repeats of three amino acid units -(Gly-X-Y)$_n$- where Gly is a glycine and X and Y can be any other amino acid but often, in mammals, X is a proline and Y is a hydroxyproline (2). The collagen molecule has propensity to self-associate into fibrils that further organize into hierarchical architectures both *in vivo* and *in vitro* (3, 4). Within those units, collagen molecules are arranged longitudinally and laterally so as forming a crystal-like period called “D period” of 67 nm evidenced both by X-ray diffraction (XRD) and electron microscopy studies (5,6). So far, the

highest resolution obtained on the whole molecule by XRD is 5-11 Å (lateral and axial resolution respectively) (5). High-resolution local structural data were obtained on several crystals of collagen-like peptide. However, due to the limited number of these crystals (~20 resolved structures currently in the PDB) data collected are far from representing the great diversity of -(Gly-X-Y)$_n$- sequences present in collagen I (7). ^{1}H NMR has only shown moderate potential to solve collagen-like peptide structure because of many overlaps of the resonances (8). Moreover, up to now ^{13}C NMR experiments did not display a sufficient spectral resolution to allow a clear distinction among the different amino acid carbon resonances (9, 10). NMR studies using specific residue labeling have given valuable information on dynamics (11, 12, 13). In order to deduce the backbone conformation, Saito H. *et al* have compared ^{13}C chemical shifts of fibrillar collagen with those obtained on small triple helical peptides (9). Torchia *et al.* have compared native and denatured resonances of collagen fragment CB2 (14). More recently, Zhu *et al.* have studied structural modifications of bovine cortical bone induced by dehydration (10). Even though all these studies reveal important structural characteristics of collagen, many ^{13}C chemical shifts were "hidden" due to a poor spectral resolution.

In the present work, we have investigated, using solid-state NMR, structural changes in collagen at high concentrations in four different regimes: three native conditions defined as (*i*) soluble at acidic pH (pH 2.5; T = 20°C), (*ii*) and (*iii*) fibrillar at basic pH (pH 8.5; T = 20°C and T = 30°C), and a denatured condition (*iv*) (pH 8.5; T = 35°C). This last sample was used as a reference. We have observed spectrally resolved 1D and 2D {^{1}H}-^{13}C MAS spectra (^{13}C typical line width from 20 to 40 Hz). We have identified side chains and backbone ^{13}C chemical shifts modifications upon fribillogenesis passing from acidic to basic pH. We have shown that one given residue can adopt different conformations along the molecule. We have interpreted our data in terms of side chain conformational modifications using statistical data from ref (15). We have also quantified those structural modifications by comparing the ^{13}C chemical shifts of denatured collagen to side chain structural library data.

EXPERIMENTAL PROCEDURES

Sample purification and fibrillogenesis

Collagen was prepared as already described (6). Collagen concentration was assessed by hydroxyproline titration (16) and purity was determined by SDS-PAGE electrophoresis. Collagen stock solution (500 mM CH_3COOH, pH=2.5) was dialyzed against HCl (3 mM, pH=2.5) and concentrated to 410 mg.mL^{-1} by a centrifugation-filtration process using PES tubes of 30 KDa cut-off (3000 g at 10°C, variable angle centrifuge EPENDORF 5702R™). The pH was further increased by exposing the acidic sample to ammonia vapors (30 w %) at 20°C for 7 minutes. 2D solid state NMR spectra were recorded 6 hours after the sample exposition to ammonia and acquired during 6 hours. The stability of the sample was checked by recording 1D ^{13}C spectra prior and after each 2D ^{1}H-^{13}C experiments. The spectra were similar for each condition. Finally, the "basic-pH" sample was centrifuged-filtrated in order to collect enough liquid (~150 µl) to measure the pH with a pH paper. The resulting pH was 8.5 ± 0.5. Denaturation of the sample was obtained by exposing the sample to a high power ^{1}H decoupling ($\nu(^1H)$ = 60 kHz) for 16 hours.

NMR parameters

Solid-state NMR analyses were performed on a Bruker Avance III 500 spectrometer at 11.7 T operating at 500.21 MHz for ^{1}H and 125.79 MHz for ^{13}C. A Bruker Magic Angle Spinning (MAS) double resonance probe was used with 7 mm rotors spun at a frequency ν_{MAS} = 2.5 kHz A BCU-X unit was used to regulate the temperature between 20°C and 35°C. For all experiments, $t_{90}(^1H)$ and $t_{90}(^{13}C)$ were 11 and 7 µs respectively. Relaxation delays of 1 and 3 s for ^{1}H and ^{13}C respectively were sufficiently long to record quantitative spectra. 512 scans were acquired for ^{13}C single pulse-experiment, 48 scans for the 2D {^{1}H}-^{13}C INEPT MAS experiment. For all ^{13}C spectra a low power proton decoupling (4.5 kHz) was applied during acquisition. The INEPT evolution delay (1.7 ms) was matched to an averaged $^1J_{CH}$ value of 145 Hz and the refocusing delay (1.1ms) was matched in order to get the CH, CH_2 and CH_3 resonances similarly phased. 256 t_1 transients were collected for 2D {^{1}H}-^{13}C INEPT spectra. ^{1}H and ^{13}C chemical shifts were referenced (δ = 0 ppm) to TMS. No ^{13}C signal was detected through a cross polarization (CP) MAS

experiment (contact time of 10ms, 440 scans and a ^{1}H decoupling of 40kHz). See SI section figure 2.

Assignment and fit

^{1}H and ^{13}C chemical shifts assignments were based on 2D $\{^{1}H\}$-^{13}C INEPT MAS spectra (Fig. 1, Fig. SI-1), and on previous works on a collagen-like peptide (8) and native collagen I (9, 10, 17). No ambiguity was found for carbon resonances except for some Cα and Cβ for which we observed overlaps in the 2D $\{^{1}H\}$-^{13}C INEPT MAS spectra obtained in native conditions. In our assignments, we did not consider tryptophans and cysteines since collagen I from rat tendon present neither (1). We were not able to assign any resonance to methionine, histidine and tyrosine due to their very small amounts (less than 0.6% of the total amino acids content). We were not able to assign 3-hydroxyprolines (~5% of total hydroxyprolines content (18)) since such modifications shift the carbon resonances to very high downfield regions (19). This point together with the lack of data in the literature makes the assignment unreliable.

1D ^{13}C spectra were fitted using Dmfit software (20). This allows the determination of the number of underlying resonances within a given peak and, consequently, allows the extraction of either the number of residues or the number of conformations for a given carbon. A direct quantification was not possible in the specific case of Arg Cγ and Pro Cγ, which overlapped in the 1D ^{13}C spectrum (Fig. SI 2). However, we used Arg Cδ peaks, free of overlap in the 1D spectra, to deduce the relative intra-residual intensity of Arg Cγ (Table SI-3 and SI-Calculation). Since there are two times less arigine than proline, we considered arginine intensity twice as low as that of proline in the quantitative spectrum. We then subtracted this intensity to the overlap peak to deduce the intensity of Pro Cγ. Dmfit software generated standard deviations (SD) based on the match between experimental data and peak simulation. Table SI-1 shows that SDs for these fits spread from 1% to 10% depending on carbon type and physico-chemical condition. Noteworthy, in our case, a change of 0.01 to 0.02 ppm in peak position always induced an increase of the standard deviation of about 1 unit or greater. We thus estimated our resolution limit to 0.02 ppm. We were not able to fit correctly the α and some β carbons, however all linear side chain carbons have been fitted (SD always inferior to 6% at pH 8.5).

RESULTS

Dense collagen I matrices were synthesized through a two-step procedure making use of the lyotropic properties of collagen molecules at acidic pH and self-assembling properties at neutral to basic pH. Since it has been previously observed that collagen molecules assume classical fibrillar staggered in both conditions (6), we used solid state NMR to analyze structural modifications of collagen molecules in four different regimes: (*i*) soluble (pH 2.5) at 20°C, (*ii*) and (*iii*) fibrillar (pH 8.5) at 20°C and 30°C, and (*iv*) denatured (pH 8.5) at 35°C. In (*iv*), ^{13}C chemical shifts only reflect the primary structure of the collagen protein and are not subservient to higher order structural constraints (triple helix or fibrillar) (14). Therefore (*iv*) was used as a reference condition. The ^{13}C 1D and 1D $\{^{1}H\}$-^{13}C INEPT MAS spectra of a fibrillar sample are depicted in Fig. SI-2. These spectra present a remarkable spectral resolution (full width at half maximum FWHM ~ 0.2 to 0.4 ppm) compared to previous ^{13}C NMR spectra reported in the literature (9, 10, 21). The ^{13}C chemical shifts obtained from the quantitative spectrum and those obtained from the 1D $\{^{1}H\}$-^{13}C INEPT spectrum are identical (Fig. SI-2). However, no signal was detected from a 1H CP MAS experiment (Fig. SI-2), meaning that all carbons from the sample exhibit fast dynamics. To analyze the impact of pH and temperature upon the collagen structure, we used a 2D $\{^{1}H\}$-^{13}C INEPT MAS experiment to correlate the ^{1}H to the ^{13}C chemical shifts (Fig. SI-3). ^{1}H and ^{13}C chemical shifts of the denatured sample are in very good agreement with previous studies of collagen-like peptide (8, 14). In the denatured state (*iv*), most of the side-chain and backbone carbons display only one single resonance (Fig. 1, SI-1 bottom row). Backbone and side chains ^{13}C chemical shifts found in the present work are in good agreement with data extracted from a random coiled proteins library (SD of 0.07 after calibration (22, 23)). Considering all native conditions whether in acidic or neutral-basic pH, the most striking difference from (*iv*) is a split of resonances in the carbon dimension for all detectable side-chain chemical shifts. At acidic pH, although this split is less pronounced, peak fitting indicates clearly that each side-chain is composed of two to five peaks, depending on the residue (Fig. 1, SI-1 second row from bottom).

Interestingly, increasing the pH induces chemical shift dispersion for a same carbon (Fig. 1, SI-1 third row from bottom). For some carbons, the peak splitting is further increased by increasing temperature from 20°C to 30°C (Fig. 1, SI-1 top row). In native conditions, most α carbons are composed of more than one peak but, because of the poor α carbon spectral resolution, we were not able to make a proper fit of these peaks. Thus, a change of pH and/or temperature does induce changes in carbon chemical shift for α carbons although we cannot quantify it. For all native conditions Gly Cα displays two resonances centered at 43.37 ppm and 42.60 ppm, with the same relative intensity and at quite close position compared to the denatured state (*iv*) (Fig. SI-1 and Fig. SI-4). In the proton dimension, the Gly Hα less-intense peak presents a clear upfield shift (Δδ = -0.1 ppm) compared to (*iv*) (Fig. SI-1, Fig. SI-4). Considering the native conditions, the only α carbon that presents a measurable change in chemical shift in the carbon dimension is Asp Cα, for which an increase in pH induces a clear downfield shift of Δδ = +0.15 ppm (Fig. SI-1). This shift difference is in agreement with that expected considering deprotonation effects (24, 25). Amongst β carbons, Asp Cβ shows an important downfield shift of +0.6 ppm with the pH increase and, at 30°C / pH 8.5 ± 0.5, it is composed of four peaks (Fig. 1C). Ala Cβ presents two peaks at acidic pH and increasing pH induces the apparition of one less intense peak at a downfield position (Fig 1A and Table SI-3). For β carbons possessing a hydroxyl group (Ser, Thr), no significant changes are seen.

In order to better describe the effects of pH and temperature, we have divided the side-chain carbons into three classes based on similarity of changes observed in carbon chemical shift. The *first class* corresponds to Arg, Lys, Glu and Gln residues. In native conditions, side chain carbons (Cγ, Cδ, and Cε) of these residues present three peaks in the carbon dimension. The central resonance is always very close to that of the denatured state (*iv*) and the two other peaks are located upfield and downfield. The increase in pH and temperature does not influence much the middle peak position but induces a further shift of the two others (Fig. 1, SI-1 second row from top). The *second class* contains only Pro and Hyp residues. The increase in pH and temperature induces an important (Δδ ~ -0.6 ppm) upfield shift of the proline Cβ and Cγ residues (~40% of the total intensity) (Fig. 1B). Hyp Cβ presents mainly three peaks at acidic pH and mainly five peaks at pH 8.5±0.5 both at 20°C and 30°C (Fig. 1C). No significant changes are seen for Hyp Cγ (Fig. SI-1) similarly to other carbons possessing hydroxyl groups. We were not able to fit correctly Cδ resonance because of peak overlap. Hyp Cδ and Pro Cδ show larger line widths than the other prolyl carbons making measurements of the less intense peaks difficult. In the *third class*, we find Leu (Cδ 1 and 2), Ile (Cδ 1), Val (Cγ 1 and 2) and Thr (Cγ) residues. The pH increase induces a split in the carbon dimension while increasing temperature does not induce important changes except for Leu (Cδ 1 and 2) and Thr (Cγ) that present less intense peaks and show a slight increase in chemical shift dispersion (Fig. 1A).

The effect of physico-chemical conditions on the resonance dispersion of Argine, Proline Cγ and Hydroxyproline Cβ is shown on Fig. SI-5. Furthermore, Fig. 2A-C compares quantitatively the increases in ^{13}C chemical shift dispersion for these residues. The graphs are normalized using chemical shift values from (*iv*). In particular, Fig.2A reveals that, for Arg and Lys, pH and temperature have a greater impact on the most upfield peaks (Δδ = +0.75 ppm) than on the more downfield ones (Δδ = -0.17 ppm). In contrast, Gln exhibits a similar evolution of the chemical shift dispersion for the two extreme peaks (Δδ = ±0.3 ppm). Concerning Glu Cγ peaks, the pH increase induces an important downfield shift (Δδ=±1.65 ppm). However one cannot directly compare acidic condition with (*iv*) (pH 8.5±0.5) due to the deprotonations occurring at basic pH. For the *second class*, Hyp Cβ and Pro Cβ and Pro Cγ, only part of the peaks shows an upfield shift (Fig. 2B). Increase in temperature induces an important shift of the resonances for Hyp and Pro (Fig. 2B, Fig. SI-5B,C). Since our 1D ^{13}C experiments were made under quantitative conditions, it is worthwhile to compare the signal intensities between the 2D INEPT experiment (sensitive to local motion) and the 1D quantitative one. The difference between 2D and 1D quantitative experiment is relatively small (<20%) (Figure SI-2, Table SI-3).

DISCUSSION

Collagen dynamics

The spectra obtained through 2D {^{1}H}-^{13}C INEPT MAS and 1D ^{13}C experiments are in good agreement with previous ^{13}C NMR experiments reported on native and denatured samples (14). We observe that varying the pH does not impact significantly the ^{1}H line widths for the native conditions indicating that the structural changes induced by fibrillogenesis affect only moderately collagen dynamics reflected by the line width of the ^{1}H resonances (Table SI-1). Mosser et al. have shown that increasing the pH of a very dense collagen solution (400 – 800 mg.mL^{-1}) resulted in the formation of very small fibrils (around 5-15 nm of diameter) (26). Nano-fibrils are also present within larger collagen I fibrils (5). As in this study we are dealing with very small fibrils subject to fast dynamics, this explains why there is no strong difference in resolution, or in INEPT signal intensity, between the soluble state (acidic pH) and the fibrillar one. Moreover, this also explains the lack of spectral resolution present in former studies achieved on Achilles tendons. In this tissue, the fibrils display a diameter at least five times larger in which the nanofibrils must be more constrained (ex. 50 nm to 200 nm in ref. 27). This also explains why we were unable to record any ^{13}C signal through a CP MAS experiment and the very strong signal of the INEPT experiments for all carbons.

Considering side-chain carbons (γ, δ, ε), for one given side chain type, our data show that the relative distribution of intensity in the 2D INEPT spectra is always close to that found in 1D quantitative spectra (Table SI-3). This result is consistent with the fact that in hydrated fibrils, there is no strong interaction between amino acid residues (11-13). In a future work a more quantitative analysis will be made concerning collagen dynamics. In the present work, we have chosen to focus our study on the structural changes revealed by the chemical shift analysis.

Chemical shift analysis

Previous works have shown that the chemical shifts of Cα, Cβ and the overall side-chain carbons (Cγ, Cδ, Cε), are sensitive to changes in backbone (ψ, ϕ) and side chains (χ_1, χ_2, χ_3) dihedral angles, respectively. In contrast, it is considered that chemical shifts of C=O carbons or protons, are sensitive to hydrogen bond and hydration degree (15, 24, 28, 29). The physical origin of the correlation between chemical shift and residue conformation is known as "γ-effect": some side chain conformations induce a *gauche* interaction for a given carbon that increases steric conflict and generates an upfield ^{13}C chemical shift (Fig. SI-6) (15). For proteins, this effect overcomes all other effects such as the ring currents (29). Since at room temperature, amino acid side chains of rat tendon fibrils show a very fast inter-conversion rate between different conformations (11-13) each ^{13}C chemical shift in our spectra must represent a equilibrium between those conformations in a specific locations along the polypeptide chain. In the denatured condition (*iv*), when no specific protein-fold is present, almost all carbons display only one ^{13}C chemical shift per carbon type and thus only a single equilibrium exists for all amino acids wherever their location in the primary sequence. On the contrary, in the triple helix and in the fibrillar state, conformations will be stabilized or destabilized according to the amino acid location in the molecule thus generating different equilibria. For example, in our case that deals with the full collagen triple helix, Arg Cγ ^{13}C resonances display the presence of three different chemical shifts indicating the presence of three major equilibria along the molecule. Moreover, the work of Dunbrack et al. (30) shows that there is strong correlation between dihedral angles inside a side chains and between side chains and backbone dihedral angles. Therefore, any change in chemical shift for a given side chain carbon (Arg Cγ for example) must induce some changes in the chemical shift of other carbons (Arg Cδ) of the same residue. Indeed, one sees that all residues always display the same number of resonances for each of their carbons (three for Lys Cε, Cδ and Cγ). Moreover, this indicates that, at least for residues having a carbon free of overlap in the 1D quantitative experiment, all the resonances are detected in the INEPT experiments.

Backbone conformation

An extended study has been conducted to investigate backbone conformation based on sequence analysis and collagen peptide-like crystal structure (7). In particular, Bella (7) has predicted the presence of 6 major helical conformations in the human collagen I triple helix. This work shows that the poorer in iminoacids a region is, the less important the triple helical twist (7). Alanine Cβ is the most useful "probe" to detect backbone conformations due to its abundance in collagen (11%). Our

results show that, at acidic pH, Ala Cβ displays four resonances representing five different backbone conformations. However, in fibrillar state, Ala Cβ chemical shifts indicates that five major backbone conformations exist in agreement with prediction of ref (7) (Fig. 1 A and Table SI-2). Ala Cβ peak at 16.6 ppm is not affected by the association into triple helix since this chemical shift remains comparable to that of (*iv*) (Fig. 1 A, Table SI-2). In random coiled peptides, residues immediately followed by a proline present a chemical upfield shift of ~1 ppm for Cα and Cβ carbons (23). Since the peak fitting for (*iv*) (Table SI-2) agrees reasonably well with the number of Ala followed by an iminoacid in the primary sequence (1), we assigned this peak (16.6 ppm) to a region rich in (Gly-Ala-Hyp) triplets. At this point, the others peaks, 17.14, 17.52, 17.85 ppm, correspond to regions rather depleted in iminoacids and the 18.2 ppm peaks correspond to the richest region.

Side chain analysis: Equilibrium quantification

In order to interpret our chemical shifts in terms of conformational changes, we have compared our data with the work of London *et al.* (15) in which the authors have built a library that correlates side chain conformations found in the PDB with ^{13}C chemical shifts. In this library, chemical shifts are those of proteins in solution at room temperature and, because of rotamer averaging, they are much lower than chemical shifts obtained for pure conformers (15). Using this library's data we were able to associate a given upfield or downfield shift with a given conformer stabilization. For example, the library shows for glutamine (Gln) that, statistically, when the side chain dihedral angle χ_1 adopts the *trans* (180°) or *–gauche* (60°) conformation, the Gln Cγ chemical shift is about 0.5 ppm lower than when χ_1 assumes the *+gauche* (-60°) conformation (15). Based on these data, a downfield shift (higher ppm value) for Gln Cγ means a stabilization of the *+gauche* conformation for χ_1. Hansen *et al.* (29) have obtained chemical shift differences between pure conformations using the J coupling constant and estimated a chemical shift difference of 5.5 ppm between Ile $C\delta_1$ conformers. These calculated chemical shift differences are always close to the difference found between the two more extreme values of the London's library (15). To extract structural information from our data, we made the assumption that the chemical shift values displayed for the denatured sample (*iv*) are close to the statistical conformers distribution found in conformation libraries (30). Indeed, as we have argued before, the lack of higher structural order in the denatured state liberates backbone and side chains into "random coiled" conformations. Then, the conformation equilibrium between side chain conformations (*+gauche*, *-gauche* and *trans*) only depends on the residue. For example, this explains why in (*iv*), all arginines present the same ^{13}C chemical shift. Thus, the chemical shift values from (*iv*) are in agreement with chemical shift values in the random coiled peptide side-chain library (23) and with denatured collagen (14) found in the literature, even though the primary sequence and the physico-chemical conditions are not the same (14, 23). We thus used ^{13}C chemical shifts values from (*iv*) as internal references of "random coiled side chain chemical shift" and have directly associated them with a "random coiled" side chain statistical distribution using an appropriate statistical structural library (30). We also made the assumption that the amplitude of the chemical shift differences between two pure conformations could be deduced from the subtraction of the more extreme chemical shift values in London's library (15). Detailed calculations with glutamine, glutamate, arginine, lysine, leucine and lysine are given as examples in the supporting information section (SI-Calculation and dihedral angle analysis). Thanks to the calculations summarized in Table 1, we show that some Arg and Lys residues adopt a (g+).(t) conformation. This conformation is characterised by a side chain location that is extended away from the triple helix (Figure SI-7). It also shows that fibrillogeneseis has an impact on Leu residues, half of them being stabilized in an extended (g+).(t) conformation and the other in a (t).(g-). It is important to know how many residues are implicated in those conformations. For glutamine, glutamate and leucine, this information is directly accessible from 1D quantitative spectrum (Table SI-3). In the case of Arg and Lys, the strong correlation between dihedral angle conformations (30) makes possible a safe estimation (SI-Calculation and dihedral angle analysis). Therefore, from 1D 13C peak quantification analysis (Table SI-3), one can see that pH and temperature changes affect many residues (about 1/3 of the total glutamine, glutamate, arginine, lysine residues).

Iminoacids

Since proline and hydroxyproline play an important role in collagen structure, they have been greatly investigated in the past (31, 32, 33). In native collagen I, as backbone dihedral angle conformations are always *trans* (8), there are only two possible stable conformations for iminoacids: pucker up and down. Actually, since the prolyl ring structure is rigid, any change in one dihedral angle affects all the others (8). It has been shown that there is a clear correlation between iminoacid positioning within the Gly-X-Y collagen triplet and pucker conformation. Residues in X and Y positions adopt preferentially a down and up conformation respectively (2). However, our data show that Cβ, Cγ prolines residues displayed an important downfield shift that increases (~0.6 ppm for β, γ) with increasing pH, suggesting that fibrillogenesis strongly influences proline conformation (Table SI-1, Table SI-3). According to statistical data, a downfield shift for Proline Cγ means a stabilization of the χ_1 angle in (g+) conformation indicating an up-pucker conformation. Thus, our data suggest that fibrillogenesis favors the up-pucker conformation. Okuyama *et al*. (28) have shown that at room temperature the crystal structure of -(Gly-Pro-Hyp)n- displays all prolines in a down-conformation, but the same peptide at low temperature (100 K) displays nearly half of all prolines (3 out of 7) in a up-conformation. Our data show that the chemical shift variation of Cδ is about half (~0.3 ppm) of that detected for Cβ and Cγ. This is in agreement with the simulated relative physical amplitude for these two carbons between down and up conformations (31). Peak quantification from 1D ^{13}C spectra reveals that about 25% of prolines undergo this stabilization.

In the case of hydroxyproline, our results show an increase of the Hyp Cβ chemical shift dispersion as the pH and temperature increase. On the contrary, Hyp Cγ does not display any significant dispersion. This is in agreement with the fact that ^{13}C chemical shift of carbons bonded to hydroxyl groups is governed by the strength of the hydrogen bond (19). For Hyp Cβ a direct quantification can be made from the 1D spectrum indicating how many residues are implicated in these conformational changes. Considering the same arguments for Pro residues, Hyp Cβ chemical shift analysis indicates that fibril formation stabilizes some hydroxyprolines in the up pucker conformation (~20%) and others in the down (~20%) (Table SI-3). This is the first time that this Pro and Hyp pucker conformational change is evidenced in the collagen I molecule. This effect seems to be essential for fibril formation since it affects a great amount of iminoacids which all together represent ~22% of collagen amino acids (1).

Impact on collagen fibrillogenesis

The fact that there are a limited number of side chains equilibria (three to five) per amino acid is a very important point to understand collagen fibrillogenesis. Indeed one sees that in the quaternary-fibrillar structure model (5), a given amino acid type from one triple helix will "face" a great diversity of amino acids of the neighbouring triple helices. However, our data show that these numerous combinations of neighbouring amino acids in the quaternary structure induce only a restricted number of side chain conformation. The present data also suggest that the relatively weak temperature-induced-attractive force component responsible for fibrillogenesis comes from rather large proportion of amino acids of the molecule.

Actually, previous force measurement studies, at neutral pH, demonstrated that collagen's self-assembling net force comes mainly from hydration forces (34). The same work indicates that the attractive component of this net force is strongly temperature dependent, in contrast to the repulsive, "hard-wall"-like component. To explain this temperature dependence, Leikin *et al.* hypothesize that this attractive component is due to reorientation of some side chains improving complementarities between the water-layer surfaces of adjacent molecules. Since this attractive component is very weak (~3.5 kcal/mol per molecule), they suggest that those side chains must be located in histidine rich regions that represent less than 1% of the whole molecular surface. On the contrary, according to their model, the repulsive forces should come from the hydration layer of the non-histidine rich regions (~99% of collagen surface).

In our case, increasing the pH "turns off" the repulsive electrostatic force making the molecules come close together to be submitted to this water layer "hard wall" repulsion. Two hypotheses can be made concerning the local level of this mechanism: -1- there is no interpenetration of water layers and/or -2- there is an energetically unfavourable interpenetration inducing a reorganization of the water shell. The first hypothesis could fit with side chains that do

not display important changes in equilibrium conformation with pH increase (the resonances having a similar chemical shift to those found in the denatured state *i.e.* ~1/2 of all side chains). In the second hypothesis, one would expect that any modifications concerning the water shell and the side-chains / side-chains or side-chains / backbone interactions would automatically induce a change in side-chain conformation. Our data show that the pH increase induces a structural reorganization for some side chains indicating that they belong to regions where water shell interpenetration has occurred. Consequently, our data also show that for some of these side-chains the increase in temperature increases the equilibrium dispersion from random coiled values. As the inter-molecular repulsive force is not affected by temperature, all side-chains showing an increase in equilibrium dispersion with both temperature and pH, must necessarily be located in regions where energetically-favoured-complementary-surfaces are formed. Moreover, considering all side chains answering this criteria (equilibrium changes with T° and pH), one can estimate that these correspond to 10% to 15% of the total collagen residue content (Table S1-1). Our results bring experimental evidences supporting the hypothesis of Leikin *et al*. (35) but further show that the proportion of residues concerned is much greater than previously postulated.

CONCLUSION

^{13}C and ^{1}H-^{13}C solid state MAS NMR experiments were used to identify conformational changes in collagen I in 3 different structural states: denatured, native-repulsive regime and native-attractive regime (fibril state). Chemical shift analysis indicates that fibrillar packing induces an increase in heterogeneity of conformations. More specifically, using Ala Cβ chemical shifts as probe, we show the existence of five different backbone conformations in collagen, one of those being specific to the fibrillar state. Analysis of ^{13}C chemical shifts demonstrates that linear side chains show moderate changes in the average conformation. Using statistical structural data present in literature, we assigned changes in ^{13}C chemical shift to the most spread-out side chain conformations and thus calculated equlibrium distributions between conformations. We also show that fibrils formation impacts on many iminoacids, stabilising prolines in the down pucker conformation and hydroxyproline both in up and down. These data suggest that the relatively weak temperature-induced-attractive force component responsible for fibrillogenesis comes from large percentage of amino acids contrary to what was thought so far. Finally, our data bring a more precise view about the impact of fibrillogenesis on local collagen conformations and open wide perspectives to the study of collagen I in interaction with other extra-cellular matrix molecules.

ACKNOWLEDGEMENTS

The authors thank Olivier Lequin for fruitful discussions at an early stage of manuscript preparation, the reviewers and Sandrine Quignard for reading and revising the manuscript. PDSP thanks Ecole Doctorale Iviv 387 and the UPMC for financial support.

REFERENCES

1. Bolboaca S.D., Jäntschi L. (2007) Amino acids sequence analysis on collagen. Bulletin *Usa Samv-cn*, **63-64**, 311-316.

2. Berisio R., De Simone A., Ruggiero A., Improta R., Vitagliano L. (2009) Role of side chains in collagen triple helix stabilization and partner recognition. J Pept Sci., **15**, 131-40.

3. Giraud-Guille M-M. (1988) Twisted plywood architecture of collagen fibrils in human compact bone osteons. Calcif Tissue Int., **42**, 167-180.

4. De Sa Peixoto P., Deniset-Besseau A., Schanne-Klein M.C., Mosser G. (2011) Quantitative assessment of collagen I liquid crystal organizations: role of ionic force and acidic solvent, and evidence of new phases. Soft Matter, **7**, 11203-11210.

5. Orgel J. P. R. O., Irving T. C., Miller A., and Tim J. (2006) Microfibrillar structure of type I collagen in situ. Proc Natl Acad Sci U S A, **103**, 9001-9005.

6. Gobeaux F., Mosser G., Anglo A., Panine P., Davidson P., Giraud-Guille M., Belamie E., (2008) Fibrillogenesis in dense collagen solutions : a physicochemical study. J Mol Biol. **376**, 1509-1522.

7. Bella J. (2010) A new method for describing the helical conformation of collagen: Dependence of the triple helical twist on amino acid sequence. J Struct Biol, **170**, 377-391.

8. Li M., Fan P., Brodsky B., Baum J. (1993) Two-dimensional NMR assignments and conformation of (Pro-Hyp-Gly)10 and a designed collagen triple-helical peptide. Biochemistry, **32**, 7377-7387.

9. Saitô H., Tabeta R., Shoji A., Ozaki T., Ando I., Miyata T. (1984) A High-Resolution 13C-NMR Study of Collagen like Polypeptides and Collagen Fibrils in Solid State Studied by the Cross-Polarization-Magic Angle-Spinning Method. Manifestation of Conformation-Dependent 13C Chemical Shifts and Application to Conformational Characterization. Biopolymers, **23**, 2279-2297.

10. Zhu P., Xu J., Sahar N., Morris M. D., Kohn D. H., Ramamoorthy A. (2009). Time-resolved dehydration-induced structural changes in a intact bovine cortical bone revealed by solid-state NMR spectroscopy. J Am Chem Soc., **131**, 17064-17065.

11. Batchelder L. S, Sullivan C. E, Jelinski L. W, Torchia D. A. (1982) Characterization of leucine side-chain reorientation in collagen-fibrils by solid-state ²H NMR. Proc Natl Acad Sci U S A, **79**, 386-389.

12. Jelinski L. W., Sullivan C. E. and Torchia D.A., (1980) 2H NMR study of molecular motion in collagen fibrils. Nature, **284**, 531-534.

13. Jelinski L. W., Torchia D. A. (1980) Investigation of labeled amino acid side-chain motion in collagen using 13C nuclear magnetic resonance. J Mol Biol., **138**, 255-272.

14. Torchia D. A., Lyerla J. R., Quattrone A. J. (1975) Molecular dynamics and structure of the random coil and helical states of the collagen peptide, 03b11-CB2, as determined by 13C magnetic resonance. Biochemistry, **14**, 887-900.

15. London R. E., Wingad B. D., Mueller G. A. (2008) Dependence of Amino Acid Side Chain 13C Shifts on Dihedral Angle : Application to Conformational Analysis. J Am Chem Soc., **130**, 11097-11105.

16. Bergman I., Loxley R., (1963) Improved and Simplified Methods for Spectrophotometric Determination of Hydroxyproline. Anal Chem. **35**, 1961-1965.

17. Aliev A. E. (2005) Solid-state NMR studies of collagen-based parchments and gelatin. Biopolymers, **77**, 230-245.

18. Eyre D. R., Weis M. A., Hudson D. M., Wu J-J., Kim L. (2011) A Novel 3-Hydroxyproline (3Hyp)-rich Motif Marks the Triple-helical C Terminus of Tendon Type I Collagen. J Biol Chem., **286**, 7732-7736.

19. Ejchart, A. (1977), 13C n.m.r. chemical shifts in aliphatic alcohols and the 03b3-shift caused by hydroxyl and methyl groups. Organic Magnetic Resonance, **9**, 351-354.

20. Massiot D., Fayon F., Capron M., King I., Le Calve S., Alonso B., Durand J.O., Bujoli B., Gan Z.H., Hoatson G. (2002) Modelling one- and two-dimensional Solid-State NMR spectra. Magn Reson Chem., **40**, 70-76.

21. Reichert D., Pascui 0., DeAzevedo E. R., Bonagamba T. J., Arnold K., Huster D. (2004) A solid-state NMR study of the fast and slow dynamics of collagen fibrils at varying hydration levels. Magn Reson Chem., 42, 276-284.

22. Tamiola K., Acar B., Mulder F. A. (2010) Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. J Am Chem Soc., **132**, 18000-18003.

23. Wishart D.S., Sykes B. D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR, **4**, 171-180.

24. de Dios A. C., Oldfield E. (1994) Chemical shifts of carbonyl carbons in peptides and proteins. J Am Chem Soc. **116**, 11485-11488.

25. Daisuke K., Toshie S., Fuyuhiko I. (1991) Characterization of pH titration shifts for all the nonlabile proton resonances in a protein by two-dimensional NMR: the case of mouse epidermal growth factor. Biochemistry, **30**, 4896-4900.

26. Mosser G., Anglo A., Helary C., Bouligand Y., Giraud-Guille M.-M., (2006) Dense tissue-like collagen matrices formed in cell-free conditions. Matrix Biol., **25**, 3-13.

27. Kadler K. E., Baldock C., Bella J., Boot-Handford R. P. (2007) Collagens at a glance. J Cell Sci., **120**, 1955-1958.

28. Richarz R. and Wuthrich K., (1978) Carbon-13 NMR Chemical Shifts of the Common Amino Acid Residues Measured in Aqueous Solutions of the Linear Tetrapeptides H-Gly-Gly-X-LAla-OH. Biopolymers, **17**, 2133-2141.

29. Hansen D. F., Neudecker P. and Kay L. E. (2010) Determination of Isoleucine Side-Chain Conformations in Ground and Excited States of Proteins from Chemical Shifts. J Am Chem Soc., **132**, 7589-7591.

30. Dunbrack R. L., Karplus M. (1994) Conformational analysis of the backbone-dependent rotamer preferences of proteins side chains. Nat Struct Biol., **1**, 334-340.

31. Mikhailov D., Mayo A. D. V. H., Mayo. K. H. (1995) 13C Multiplet Nuclear Magnetic Resonance Relaxation-Derived Ring Puckering and Backbone Dynamics in Proline-Containing Glycine-Based Peptides. Biophys J., **68**, 1540-1550.

32. Mooney D. S., Kollman A. P., Klein E. T. (2002) Conformational Preferences of Substituted Prolines in the Collagen Triple Helix. Biopolymers, **64**, 63-71.

33. Okuyama K., Hongo C., Fukushima R., Wu G., Narita H., Noguchi K., Tanaka Y., Nishino N. (2004) Crystal structures of collagen model peptides with Pro-Hyp-Gly repeating sequence at 1.26 A resolution: implications for proline ring puckering. Biopolymers, **76**, 367-77.

34. Leikin S., Rau D.C., Parsegian V.A. (1994) Direct Measurement of Forces Between Self-Assembled Proteins: Temperature-Dependent Exponential Forces Between Collagen Triple Helices. Proc Natl Acad Sci USA, **91**, 276-280.

35. Leikin S., Rau D. C., Parsegian V. A. (1995) Temperature-favoured assembly of collagen is driven by hydrophilic not hydrophobic interactions. Nat Struct Biol., **2**, 205-10.

36. Kramer R. Z., Bella J., Brodsky B. and Berman H. M. (2001) The Crystal and Molecular Structure of a Collagen-like Peptide With A Biologically Relevant Sequence. J Mol Biol., **311**, 131-147.

37. Han B., Liu Y., Ginzinger S., and Wishart D. (2011) SHIFTX2: significantly improved protein chemical shift prediction. J. Biomol NMR, **50**, 43-57.

FIGURE LEGENDS

Figure 1. 2D {^{1}H}-^{13}C INEPT MAS spectra of a 410 mg.mL^{-1} collagen sample. From bottom to top: (*iv*) denatured - pH 8.5, 35°C, (*i*) soluble - pH 2.5, 20°C, (*ii*) fibrillar - pH 8.5, 20°C, (*iii*) 30°C. In (*iv*) almost each individual carbon is represented by a unique peak. In native states, almost all carbon side-chains resonances are composed of more than one peak and the distance between peaks increases passing from pH 2.5 to pH 8.5. A, B, C are zooms in {^{1}H} and ^{13}C chemical shifts in different areas of the full spectra. A: ^{13}C (14 - 24.5 ppm), ^{1}H (0.6 - 1.51), B : ^{13}C (23 - 35), ^{1}H (1.35 - 2.50), C : ^{13}C (31 - 43), ^{1}H (1.90 - 3.50).

Figure 2. Chemical shifts dispersion and equilibrium constants for some side chains. Plot of the chemical shift dispersion defined as ($\delta - \delta_{den}$) with δ_{den} being the corresponding chemical shift in (*iv*) condition and d the corresponding chemical shift in conditions (*i*), (*ii*) and (*iii*) conditions. A) Arginine (Arg), lysine (Lys) and glutamine (Gln) carbons γ. For these residues we have omitted the central peaks in the graph since they were not greatly affected by pH or temperature changes.B) Hydroxyproline Cβ and Proline Cγ and Cβ. C) Leucine Cδ1, Threonine Cγ2 and Isoleucine Cγ2. D) Equilibrium constants between the conformers g+ and t/g-, for Arginine, Lysine and Glutamine χ1 dihedral angles in the different states.

TABLE LEGEND

Table 1. ^{13}C chemical shift (CS) and calculated equilibrium constant (K) for side chain carbons of collagen sample in denaturated, fibrillar and soluble conditions. Chemical shift (δ) and peak intensity were measured using Dmfit 2008 as described in the experimental section. The equilibrium constant calculation is detailed in SI. Since the state of protonation of glutamates at pH 2.5 (*i*) is different from that of the reference (*iv*) we were not able to calculate its equilibrium constant.

Figure 1

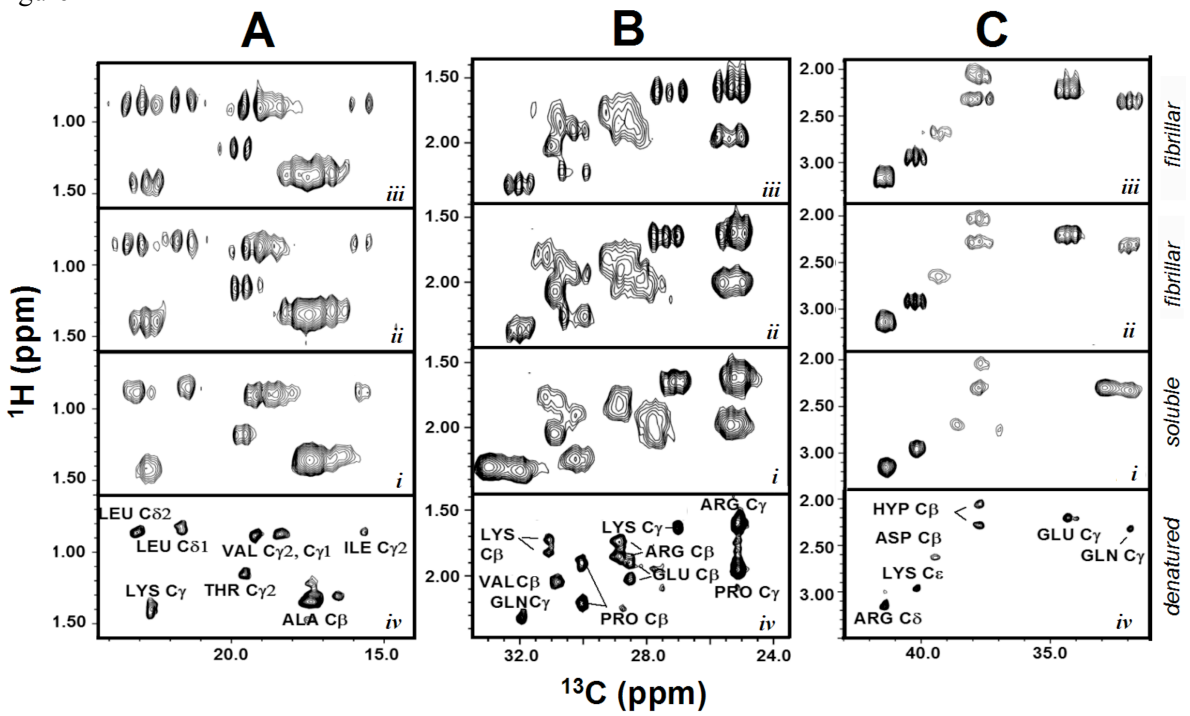
A
B
C
fibrillar
fibrillar
soluble
denatured
iii
ii
i
iv
LEU Cδ2
LEU Cδ1
VAL Cγ2, Cγ1
ILE Cγ2
LYS Cγ
THR Cγ2
ALA Cβ
LYS Cβ
LYS Cγ
ARG Cγ
ARG Cβ
GLU Cβ
VAL Cβ
GLN Cγ
PRO Cγ
PRO Cβ
HYP Cβ
ASP Cβ
LYS Cε
ARG Cδ
GLU Cγ
GLN Cγ
1H (ppm)
13C (ppm)

Figure 2

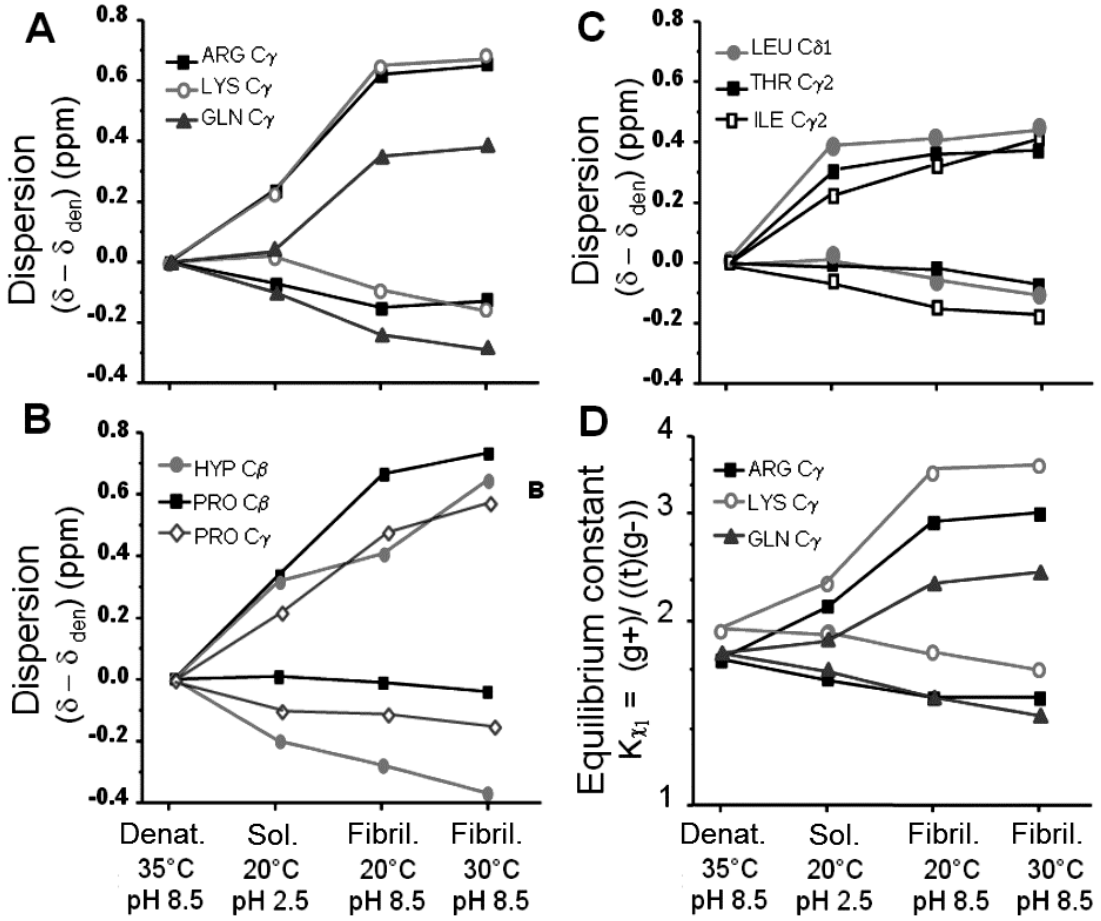
A
ARG Cγ
LYS Cγ
GLN Cγ
Dispersion (δ – δ den) (ppm)
B
HYP Cβ
PRO Cβ
PRO Cγ
C
LEU Cδ1
THR Cγ2
ILE Cγ2
D
Equilibrium constant Kχ1 = (g+)/ ((t)(g-))
Denat. 35°C pH 8.5
Sol. 20°C pH 2.5
Fibril. 20°C pH 8.5
Fibril. 30°C pH 8.5

Table 1

Table 1 :

	(*iv*) Denatured pH 8,5 35°C	(*iii*) Fibrillar pH 8,5 30°C		(*ii*) Fibrillar pH 8,5 20°C		(*i*) Soluble pH 2,5 20°C	
	δ (ppm)	δ (ppm)	K (t,g-)/(g+)	δ (ppm)	K (t,g-)/(g+)	δ (ppm)	K (t,g-)/(g+)
			(g+)/(t)		(g+)/(t)		(g+)/(t)
ARG	25,04	25,69	3,0	25,66	2,9	25,56	2,7
Cγ		25,29	2,1	25,24	2,0	25,28	2,1
		24,91	1,6	24,89	1,5	24,97	1,6
			(t)/(g+,g-)		(t)/(g+,g-)		(t)/(g+,g-)
ARG	41,37	41,72	5,8	41,70	5,6	41,52	4,4
Cδ		41,48	4,2	41,48	4,2	41,40	3,9
		41,23	3,2	41,25	3,3	41,22	3,2
			(g+)/(t)		(g+)/(t)		(g+)/(t)
GLN	31,94	32,32	2,4	32,29	2,3	31,98	1,8
Cγ		31,99	1,8	31,99	1,8	32,17	2,1
		31,65	1,4	31,70	1,5	31,84	1,7
			(g+)/(t)		(g+)/(t)		(g+)/(t)
GLU	34,30	34,73	2,2	34,69	2,1	-	-
Cγ		34,42	1,7	34,39	1,6	-	-
		34,09	1,3	34,08	1,3	-	-
			(g+,g-)/(t)		(g+,g-)/(t)		(g+,g-)/(t)
ILE	15,70	16,2	92	15,92	91	15,79	89
Cγ1		15,43	83	15,46	84	15,56	85
			(g-)/(t)		(g-)/(t)		(g-)/(t)
LEU	23,08	23,98	3,5	23,91	3,3	23,47	2,3
Cδ1		23,51	2,4	23,49	2,3	23,13	1,8
		22,96	1,6	23,00	1,6	-	-
		22,53	1,2	22,48	1,1		
			(g-)/(t)		(g-)/(t)		(g-)/(t)
LEU	21,61	20,91	3,2	20,86	3,3	21,43	2,1
Cδ2		21,33	2,3	21,33	2,3	21,69	1,8
		21,88	1,6	21,83	1,6	-	-
		22,31	1,2	22,22	1,3	-	-
			(g+)/(t)		(g+)/(t)		(g+)/(t)
LYS	22,60	23,27	3,6	23,25	3,5	23,00	2,7
Cγ		22,82	2,3	22,86	2,4	22,83	2,4
		22,44	1,7	22,51	1,8	22,62	2,0
			(t)/(g+,g-)		(t)/(g+,g-)		(t)/(g+,g-)
LYS	26,90	27,61	6,0	27,78	7,9	27,46	4,9
Cδ		27,23	3,7	27,39	4,4	27,17	3,5
		26,81	2,5	27,02	3,0	26,85	2,5
			(g-,g+)/(t)		(g-,g+)/(t)		(g-,g+)/(t)
THR	19,53	20,39	1,8	19,91	1,2	19,91	1,2
Cγ2		19,92	1,2	19,51	0,9	19,51	0,9
		19,46	0,8	19,06	0,6	19,06	0,6

Solid-state NMR Study Reveals Collagen I Structural Modifications of Amino Acid Side Chains upon Fibrillogenesis

P. De Sa Peixoto, G. Laurent, T. Azaïs, G. Mosser*.

SUPPLEMENTARY INFORMATION

Calculation of Side Chains Equilibrium Constants and dihedral angle analysis

To calculate the equilibrium constants of the amino side chains we used denatured condition chemical shifts values as internal reference of "random coiled" side chain chemical shifts. We have directly associated them with a "random coiled" side chain statistical distribution using an appropriate statistical structural library[1]. We have also estimated the amplitude difference between two pure conformations using the more extreme chemical shift values in London's library (Lon). This enabled to access a chemical shift amplitude difference ($\Delta\delta_{amp}$) between two pure conformations A and B as:

$$|Dd_{amp}^{Lon}| = |d_A^{cal} - d_B^{cal}| \qquad [1]$$

Where δ_A and δ_B are the calculated (cal) chemical shifts from two pure conformations A and B respectively. δ_A and δ_B are calculated as:

$$d_A = d_{den}^{exp} + \left(1 - pA_{den}^{lit}\right) ? Dd_{amp}^{Lon} \qquad [2]$$

Where δ^{exp}_{den} is the chemical shift of the denatured condition issue from our experiments (exp) and pA^{lit}_{den} is the statistical abundance of a given conformer "A" in a random coiled structure as deduced from literature (lit).

For example, in the case of Glutamine's (Gln) χ_1 angle, statistical data present in the literature[2] show that the ratio of side chain conformations distribution is about 0.08, 0.60 and 0.32 for *–gauche(g-), +gauche(g+)* and *trans(t)* conformers, respectively. Even though this statistic distribution was collected for crystals, it is representative of the conformational distribution of Gln side chain at room temperature (15). In other words, a Gln side chain at room temperature will be in equilibrium between these three conformations but statistically the most favourable conformation will be *trans* (g+). London's library chemical shifts shows that the greatest chemical shift difference ($\Delta\delta^{lon}_{amp}$) found is 6 ppm between *trans* and *+gauche* conformers of glutamine Cɣ. Crystallographic conformer library's shows that *–gauche* is not a very populated state, indicating that it is quite energetically unfavourable (30).

[1] which we can directly associate with a "random coiled" side chain statistical distribution using an appropriated statistical structural library

[2] For example, in the case of Glutamine's (Gln)χ_1 angle, statistical data present in literature

For a protein showing very fast side chain dynamics such as collagen (11, 12, 13), the contribution to the equilibrium of such a very rare conformer will always be very small. In this case, we proceeded as Hansen *et al.* (29) and we did not considered this rare conformer in our analysis thus giving a statistical distribution for *+gauche(g+)* and *trans(t)* of 0.64 and 0.36 respectively. A will represent the (g+) conformer and B the (t) conformer and pA_x becomes the probability 0.64. Inserting theses values as well as our experimental Cγ chemical shift in denatured condition (δ_{den}= 31.94) in equation [1] and [2] gives us $\delta_{(g+)}$ and $\delta_{(t)}$ values for pure conformers g+ and t of 34.1 and 28.1 ppm respectively.

Equations [3] to [5] are now used to determine the ratio of conformers *+gauche* and *trans* for the chemical shifts obtained from our data.

$$pA_x^{cal} = \frac{d_x^{\exp} - d_B^{cal}}{Dd_{amp}^{Lon}} \qquad [3]$$

$$pB_x^{cal} = 1 - pA_x^{cal} \qquad [4]$$

$$K_x^{cal} = \frac{pA_x^{cal}}{pB_x^{cal}} \qquad [5]$$

For example, at 30°C and pH 8.5, Glutamine (Gln) Cγ displays a resonance located at 32.32 ppm. Using this value for δ^{exp}_x in equation [3], we finally obtain the equilibrium constant $K_{32.32}^{cal}[(g+)/(t)] \approx 2.4$. In our case, this model tolerates important incertitude of the chemical shift amplitude difference between the pure conformations ($\Delta\delta_{amp}$). Indeed, if we increase our amplitude difference between conformers of about 2 ppm (making it 8 ppm instead of 6 ppm), it would result in a new equilibrium constant $K_{32.32}^{cal}[(g+)/(t)]$ of 2.2, quite close to 2.4 obtained before. These results are summarized in Table 1 and Table SI-4.

Arginine and Lysine

Statistical data show that, for Argine and Lysine residues, the most populated combinations of χ_1 and χ_2 are respectively (g+).(t) with a frequency of ~ 0.44, (t).(t) ~ 0.35, (g+).(g+) ~ 0.1, (g-).(g-) ~0.1 and (t).(g-) <0.1 (30). From our data, we can see that pH increase induces a strong stabilization of χ_1 g+ for the most downfield peak (Fig. 4A, Fig. 5). This indicates an increased stabilization of either the (g+).(t) and/or (g+).(g+) conformers. As we have already argued, a strong stabilization of a rather energetically unfavorable conformation such as (g+).(g+) is very unlikely for collagen. Thus, we conclude that this upfield shift most likely indicates an increase at the (g+).(t) conformation. This is coherent with our results on χ_2 that show a strong stabilization of *trans* conformation. This is an important result since when a

linear side chain such as Arg and Lys adopts a (g+).(t) conformation, it is extended away from the triple helix (Figure 5A): this conformation is ideal to make inter-helical interaction in contrast to a (t).(t) conformation that favours inter-strand interaction (Figure 5B) (2). It is important to know how many residues are implicated in each conformation. Cγ is the most important carbon to calculate Arg and Lys χ1, χ2 dihedral angles. Unfortunately, a direct quantification is not possible since there are overlaps in the quantitative spectrum for those signals. Nevertheless, for these amino acids, we can use Arg Cδ and Lys Cε intensity from the 1D spectrum to access the percentage of residues in each conformation. Indeed, the residues displaying a conformational equilibrium generating the middle resonances of Arg Cδ and Lys Cε carbons (that do not present a strong shift with pH and temperature, as the middle peaks of Arg Cγ and Lys Cγ do) are necessarily the same residues generating the middle peak of Arg Cγ and Lys Cγ peaks (since dihedral angles are correlated). For the two others peaks (more upfield and downfield ones) we used statistical data form literature (15, 37) to make the link between them. Therefore, we can confidently say that 1/3 of Arg and Lys residues in collagen display an important stabilization of this (g+).(t) conformation.

Leucine

The specificity of Leu is that it is composed of two γ-branched atoms (Cδ1 and Cδ2). In our data, assignment of Leucine Cδ1 and Cδ2 could be easily made based on results obtained on the denatured sample. In this case the equilibrium constant is close to a statistical distribution and the crystallographic library's data could be used to directly assign these shifts. Crystallographic library shows that (t).(g-) and (g+).(t) (for Leu Cδ2 (χ_1).(χ_2) respectively) are the two main conformers for Leu. Moreover, Okuyama *et al*. (33) have argued, using steric considerations based on collagen backbone conformations, that when Leu is followed by an hydroxyproline (as in the Gly-Leu-Hyp peptide) (g+).(t) is the only stable conformation for Leu. Our results (Fig. 1A) show that half of leucines are in this (g+).(t) conformation at pH 8.5, (K (t)/(g-) ~2 and 3.5 for χ_2, indicating a stabilisation of (g+).(t) compared to (t).(g-)). Our data also shows that the pH has an impact on small amount of leucines indicating that there are some key sites in collagen where leucine conformations are highly affected by fibrillar packing. Previous crystallographic work on collagen peptides shows that when Leu side chain adopts a (g+).(t) conformation it is more exposed to solvent (33). On the contrary, the equilibrium constant calculated for the more upfield peak of Cδ2, shows an important stabilization of (t).(g-) conformation. In this conformation Leu side chain comes in close proximity with the next strand carbonyl atom.

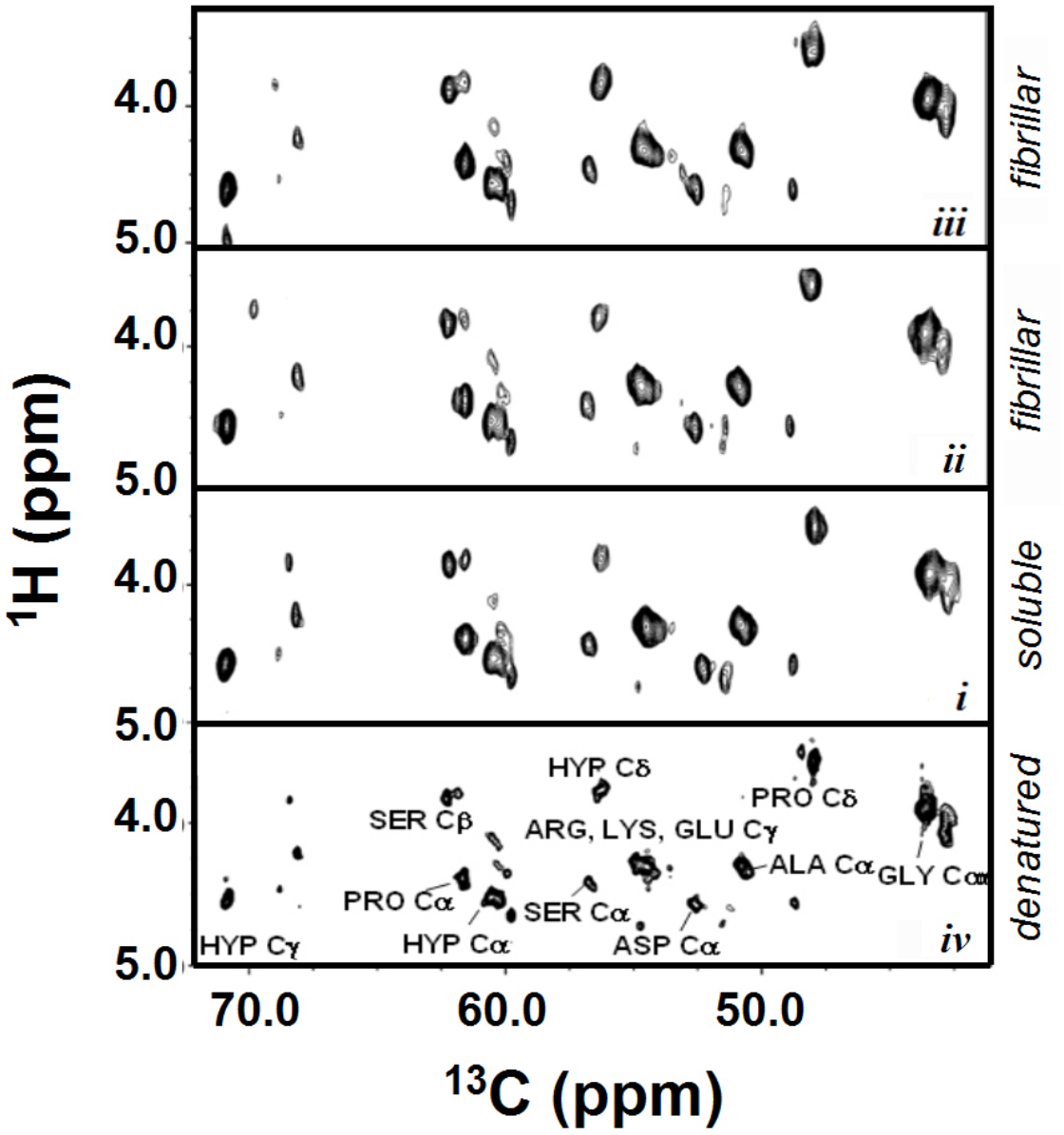

Figure SI-1. 2D {^{1}H}-^{13}C INEPT MAS spectra of a 410 mg.mL^{-1} collagen sample corresponding to the α carbon region. From bottom to top: (*iv*) denatured (pH 8.5) at 35°C, (*i*) soluble (pH 2.5) at 20°C, (*ii*) fibrillated (pH 8.5) at 20°C and (*iii*) 30°C. For α carbons, the pH modification induces less visible changes in ^{13}C chemical shifts than for side chains carbons.

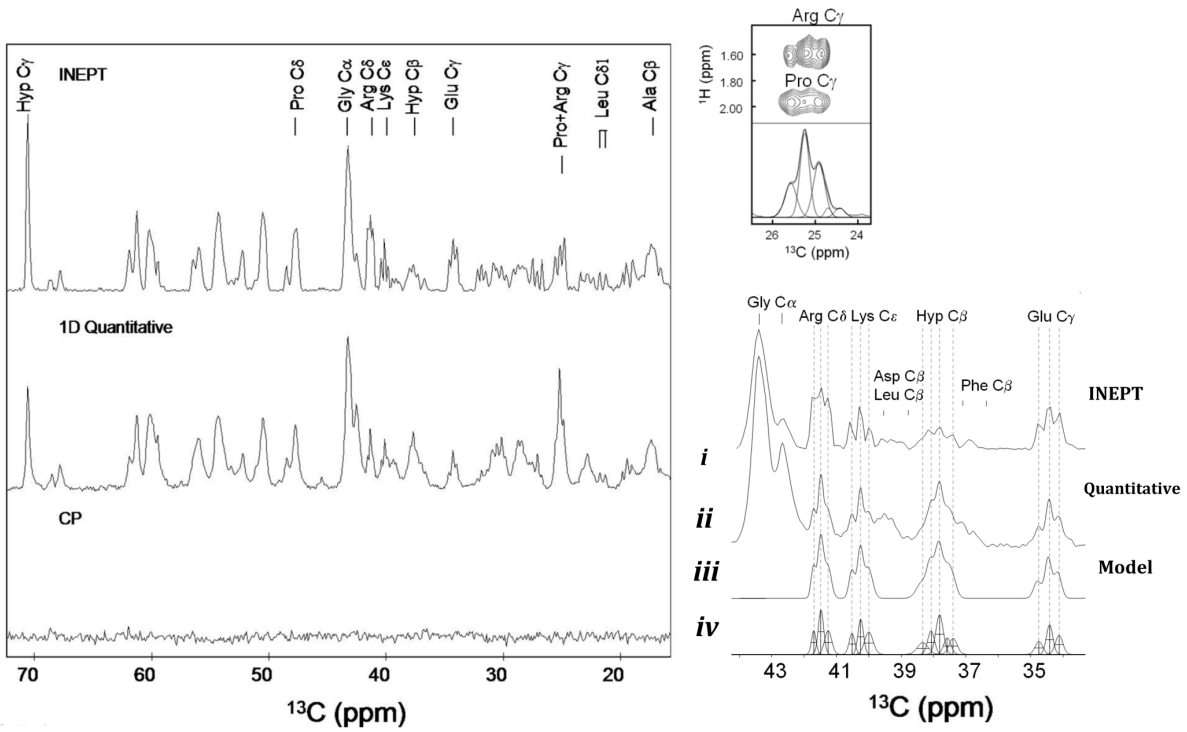

Figure SI-2: 1D MAS spectra of a 410 mg.mL^{-1} collagen sample at pH 8.5 and T = 30°C and peak quantification graph. Left, from down to top : 1D {^{1}H}-^{13}C CP MAS spectrum, 1D {^{1}H}-^{13}C quantitative spectrum and 1D {^{1}H}-^{13}C INEPT spectrum**.** The absence of signal from the CP spectrum and the narrow line widths from all spectra indicate that all residues exhibit fast motion dynamics. Right, top : Zoom at the Arg and Pro Cɣ, region for 2D and 1D quantitative spectra 30°C pH 8.5. Right, bottom : (*i*) INEPT, (*ii*) - Quantitative spectrum, (*iii*) and (*iv*) simulated spectrum of quantitative spectrum. Vertical lines show that there is a very good agreement of peak position of different side chains between INEPT and the quantitative spectrum. One can see that the relative difference in peak intensity between (*i*) and (*ii*) is not very strong. Noteworthy, in (*iii*) and (*iv*) we have extracted the other peaks from the whole simulated spectrum to improve the clarity of the figure.

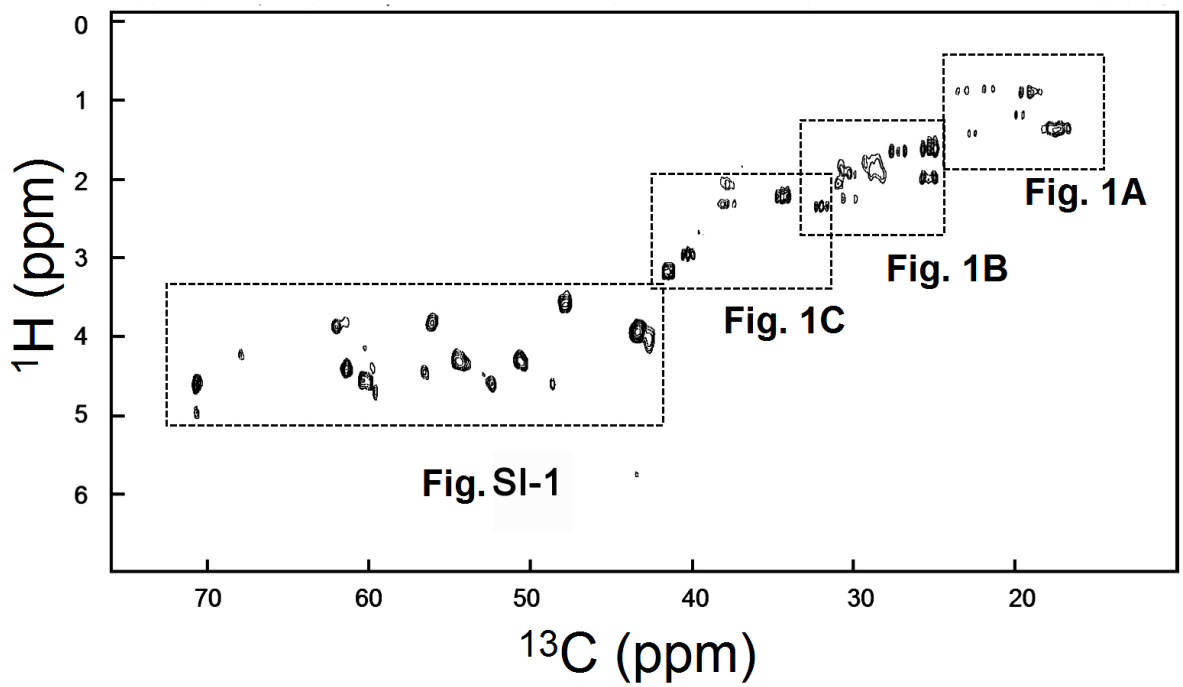

Figure SI-3: 2D {^{1}H}-^{13}C INEPT MAS spectrum of a 410 mg.mL^{-1} of a collagen sample at pH 8.5 and T = 30°C. Squares represent regions corresponding to Figure 1A, B and C and Figure SI-1.

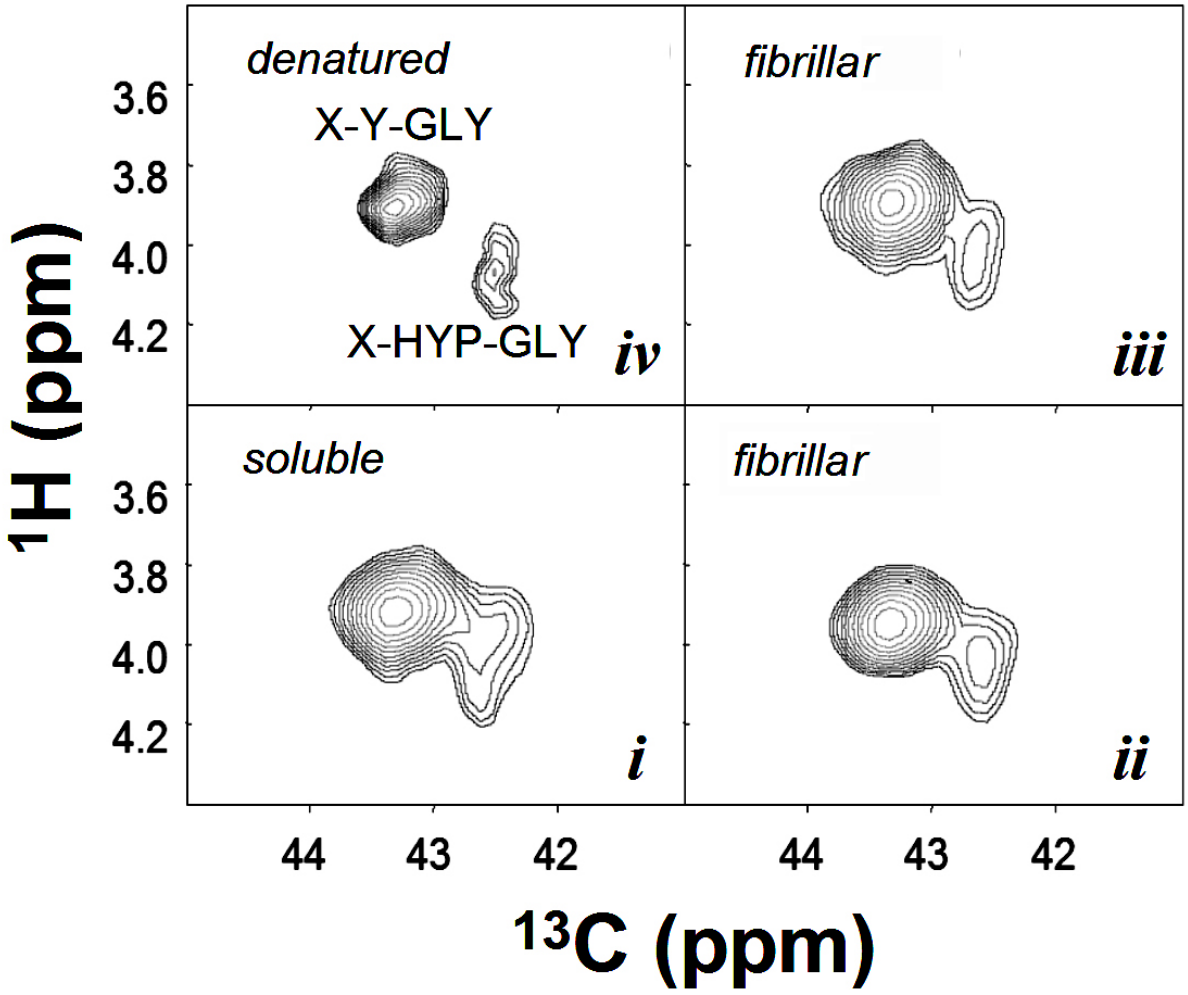

Figure SI-4: A zoom in the Glycine (Gly) region of the 2D {^{1}H}-^{13}C INEPT MAS spectra. (Top left) (*iv*) Denatured state pH 8.5, T = 35°C, (top right) (*iii*) native state pH 2.5, T = 20°C, (bottom left) (*i*) fibrillated state pH 8.5, T = 20°C and (bottom right) (*ii*) fibrillated state pH 8.5, T = 30°C.

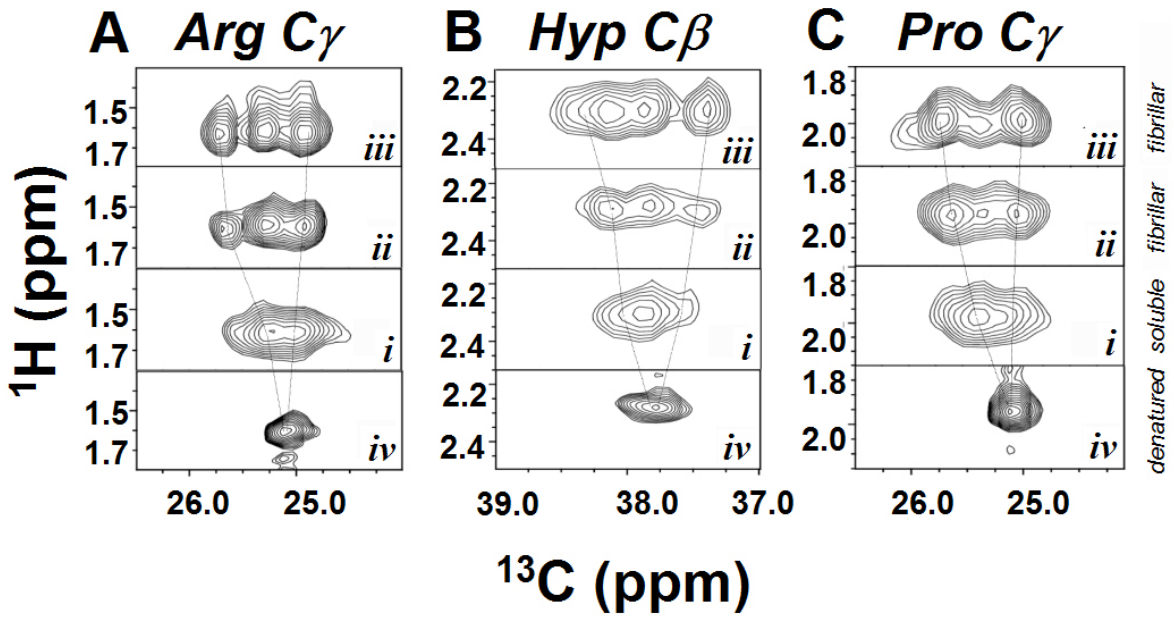

Figure SI-5. Zooms on 2D {^{1}H}-^{13}C INEPT MAS spectra. A) Arginine ɣ carbon region (Arg Cɣ), B), Hydroxyproline β carbon region (Hyp Cβ) and C) Proline ɣ carbon region (Pro Cɣ). From bottom to top: (*iv*) denatured (pH 8.5) at 35°C, (*i*) soluble (pH 2.5) at 20°C, (*iii*) fibrillated (pH 8.5) at 20°C and (*iii*) 30°C.

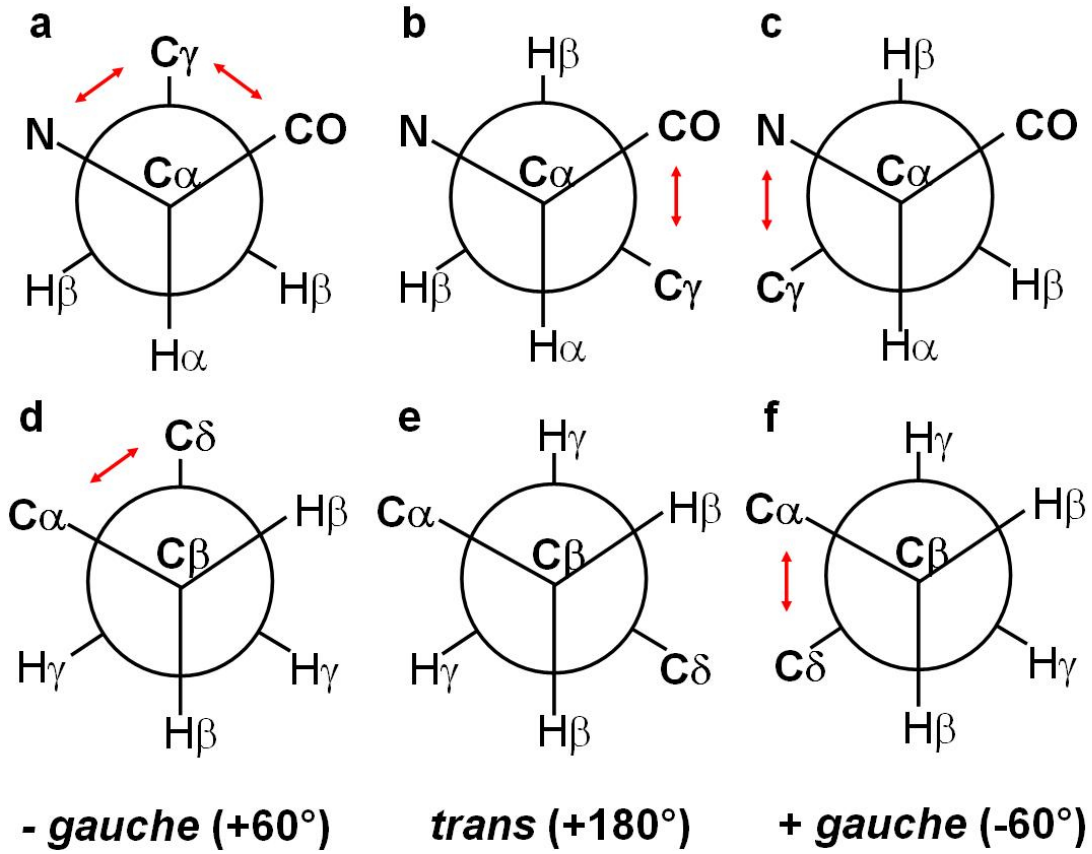

Figure SI-6. Newman projections of the three most stable conformations for χ_1 (a, b, c) and χ_2 (d, e, f). Red arrows indicate a possible steric conflict between non-hydrogen atoms. Statistical data (see reference in the text) show that the more a conformation creates steric conflicts the more the carbon presents an upfield shift.

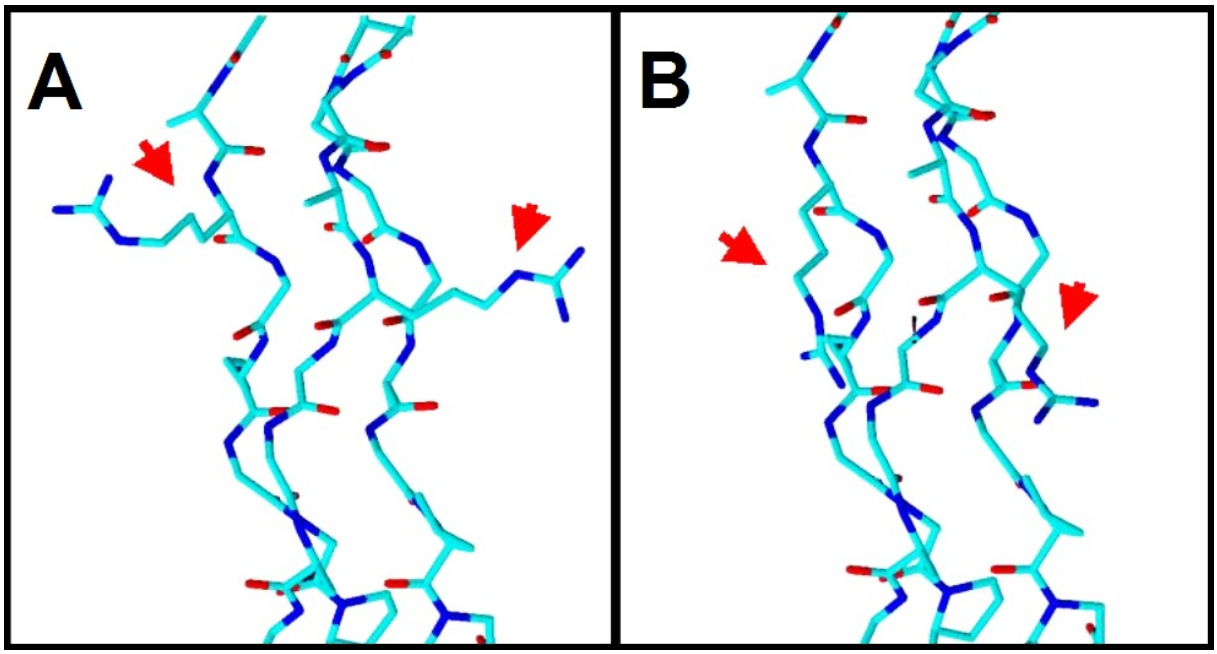

Figure SI-7. Representation of a collagen-like peptide. A), Arginines in +*gauche, trans* conformation (for χ_1 and χ_2 respectively). B) Arginines in *trans, trans* conformations (for χ_1 and χ_2 respectively). In A, side chains are considerably more extended away from the triple helix than in B which increases the probability of inter-helical interaction. The Arg dihedral angles atom's coordinates were modified from 1BKV (36). Arrows indicates Arg side chains.

Table SI-1: ^{13}C Chemical shifts (CS) and peak half width (FWHM) corresponding to different native conditions. For each condition (pH 2.5 T = 20°C, pH 8.5±0.5 T = 20°C and pH 8.5±0.5 T = 30°C) the left, middle and right column shows, respectively, the ^{13}C CS taken from 2D {^{1}H}-^{13}C spectra chemical shifts, line widths (FWHM) and the percentage of the standard deviation (SD) between the simulated model and the experimental spectrum for each group of peaks. Chemical shift (CS) and peak intensity were measured using Dmfit 2011 as described in experimental section.

Table SI-1

	(iii) Fibrillar (pH 8,5 30°C)			(ii) Fibrillar (pH 8,5 20°C)			(i) Soluble (pH 2,5 20°C)		
	δ (ppm)	FWHM (ppm)	SD (%)	δ (ppm)	FWHM (ppm)	SD (%)	δ (ppm)	FWHM (ppm)	SD (%)
ALA Cβ	18,23	0,4	4,0	18,22	0,3	6,0	17,81	0,29	8,0
	17,85	0,26		17,80	0,3		17,44	0,41	
	17,52	0,42		17,45	0,4		17,1	0,35	
	17,14	0,38		17,08	0,35		16,65	0,43	
	16,67	0,32		16,65	0,32		-	-	
ARG Cγ	25,69	0,24	2,0	25,66	0,27	2,0	25,56	0,26	2,0
	25,29	0,26		25,24	0,33		25,28	0,28	
	24,91	0,27		24,89	0,26		24,97	0,32	
ARG Cδ	41,72	0,17	1,0	41,70	0,24	1,0	41,52	0,17	2,0
	41,48	0,24		41,48	0,24		41,40	0,14	
	41,23	0,2		41,25	0,21		41,22	0,19	
ASP Cβ	39,96	0,26	1,0	39,94	0,3	3,0	38,82	0,5	6,0
	39,57	0,25		39,47	0,4		38,39	0,3	
	39,28	0,20		39,73	0,2		-	-	
	38,99	0,37		39,17	0,4		-	-	
GLN Cγ	32,32	0,27	2,0	32,29	0,32	3,0	31,98	0,21	3,0
	31,99	0,2		31,99	0,23		32,17	0,23	
	31,65	0,26		31,70	0,28		31,84	0,16	
GLU Cγ	34,73	0,21	2,5	34,69	0,33	3,5	32,98	0,58	6,0
	34,42	0,23		34,39	0,22		-	-	
	34,09	0,29		34,08	0,27		-	-	
HYP Cβ	38,40	0,32	1,0	38,16	0,24	2,5	38,07	0,32	2,0
	38,10	0,4		37,85	0,31		37,85	0,21	
	37,82	0,29		37,47	0,24		37,55	0,44	
	37,38	0,25		-	-		-	-	
ILE Cγ2	16,01	0,2	2,0	15,92	0,21	4,0	15,79	0,43	6,0
	15,41	0,24		15,46	0,22		15,56	-	
LEU Cδ1	23,98	0,28	4,5	23,91	0,21	5,0	23,47	0,29	5,0
	23,51	0,23		23,49	0,22		23,13	0,38	
	22,96	0,26		23,00	0,29		-	-	
	22,53	0,25		22,48	0,25		-	-	
LEU Cδ2	20,91	0,27	4,5	20,86	0,25	5,0	21,43	0,28	5,0
	21,33	0,24		21,33	0,27		21,69	0,24	
	21,88	0,23		21,83	0,26		-	-	
	22,31	0,2		22,22	0,28		-	-	
LYS Cγ	23,27	0,23	3,0	23,25	0,26	5,0	22,83	0,19	5,0
	22,82	0,31		22,86	0,28		22,62	0,18	
	22,44	0,24		22,51	0,25		23,00	0,17	
LYS Cδ	27,61	0,25	2,5	27,77	0,26	2,5	27,46	0,23	4,0
	27,23	0,2		27,39	0,24		27,17	0,27	
	26,81	0,2		27,02	0,25		26,85	0,27	
PRO Cβ (Hβ2)	30,66	0,21	2,0	30,59	0,25	4,0	30,24	0,35	6,0
	30,38	0,27		30,11	0,22		29,91	0,31	
	29,86	0,27		29,89	0,28		-	-	
PRO Cγ	25,70	0,24	2,0	25,60	0,36	4,0	25,34	0,37	4,0
	25,41	0,31		25,25	0,29		25,02	0,27	
	24,97	0,32		25,01	0,24		-	-	
THR Cγ2	20,39	0,22	2,5	19,91	0,2	5,0	19,91	0,23	6,0
	19,92	0,18		19,51	0,25		19,51	0,26	
	19,46	0,17		19,06	0,19		19,06	0,29	

Table SI-2: Alanine Cβ chemical shift's and relatives intensities from 1D ^{13}C and 2D {^{1}H}-^{13}C spectra at four different physico-chemical conditions (natives – pH 2.5 – T = 20°C and pH 8.5 – T = 20°C and 30°C and denatured pH 8.5 – T = 35°C). Passing from acidic pH 2.5 to pH 8.5 induces the apparition of a new peak at ~18.2 ppm and, except in denatured condition, a decrease in intensity of the middle peak at ~17.45 ppm.

Table SI-2 : ALA Cβ

	δ (ppm)	2D ^{1}H-^{13}C Integration	1D ^{13}C Integration
Fibrillar pH 8,5 30°C (*iii*)	18,23	12	13
	17,85	13	11
	17,52	30	37
	17,14	25	22
	16,67	14	16
Fibrillar pH 8,5 20°C (*ii*)	18,22	9	9
	17,80	18	17
	17,45	37	34
	17,08	18	19
	16,65	13	13
Soluble pH 2,5 20°C (*i*)	17,81	13	18
	17,44	44	48
	17,10	17	17
	16,65	13	12
Denatured pH 8,5 20°C (*iv*)	17,5	85	85
	16,58	15	15

Table SI-3: Cβ, Cγ and Cδ relatives intensities from both 1D ^{13}C and 2D {^{1}H}-^{13}C spectra at 30°C pH 8.5. Both, 1D and 2D, spectra were made on the same collagen sample (410 mg.mL^{-1}, pH 8.5 and T = 30°C). Respectively left and middle columns show carbons chemical shifts and peak integration results from the 2D {^{1}H}-^{13}C INEPT MAS spectra. The right column shows peak integration results from the 1D ^{13}C quantitative spectrum. Noteworthy, the relative peak intensities from both spectra type are always quite close indicating that the 2D spectrum presents semi quantitative intensities. *[1] Intensities estimated from Arg Cδ and Lys Cε peaks *[2] - Intensity deduced from subtraction of Arg Cδ peaks. Peak intensity was measured using Dmfit 2011 as described in experimental section.

Table SI-3 : Fibrillar pH 8.5. 30°C *(iii)*

	Chemical Shift (ppm)	2D ^{1}H-^{13}C Integration %	1D ^{13}C Integration %
ARG Cγ	25.69	25	31*[1]
	25.29	32	46*[1]
	24.91	43	22*[1]
ARG Cδ	41.72	26	22
	41.48	44	46
	41.23	28	31
GLN Cγ	32.32	34	23
	31.99	33	46
	31.65	33	-
GLU Cγ	34.73	23	23
	34.42	37	44
	34.09	40	32
HYP Cβ	38.37	10	10
	38.06	21	21
	37.76	37	37
	37.56	36	15
	37.38	24	15
ILE Cγ2	16.01	60	-
	15.41	40	-
LEU Cδ1	23.98	13	-
	23.51	29	-
	22.96	37	-
	22.53	17	-
LEU Cδ2	20.86	19	-
	21.33	33	33
	21.83	35	30
	22.22	12	-
LYS Cγ	23.27	25	29*[1]
	22.82	42	43*[1]
	22.44	33	26*[1]
LYS Cδ	27.61	44	26*[1]
	27.23	21	43*[1]
	26.81	27	29*[1]
LYS Cε	40.5	24	26
	40.24	34	43
	39.99	24	29
PRO Cβ (Hβ2)	30.66	34	-
	30.38	19	-
	29.86	34	-
PRO Cγ	25.70	28	22*[2]
	25.41	22	44*[2]
	24.97	37	34*[2]
THR Cγ2	20.39	22	-
	19.92	37	-
	19.46	36	-

Table SI-4: Values of "pure" chemical shift conformation.

Table SI-4

Chemical shifts of "pure" conformations						
		ILE Cγ2	LYS Cδ	ARG Cδ		
δ (ppm)	*trans*	16,45	28,4	42,53		
δ (ppm)	*-gauche or +gauche*	10,45	22,9	37,03		
		ARG Cγ	GLN Cγ	GLU Cγ	THR Cγ	LYS Cγ
δ (ppm)	*+gauche*	27,2	34,1	36,3	22,18	24,63
δ (ppm)	*trans or -gauche*	21,2	28,1	31,3	17,18	18,63
		LEU Cδ1	LEU Cδ2			
δ (ppm)	*-gauche*	25,3	19,35			
δ (ppm)	*trans*	18,8	25,85			