Solid-state NMR Study Reveals Collagen I Structural Modifications of Amino Acid Side Chains upon Fibrillogenesis*3

Paulo De Sa Peixoto, Guillaume Laurent, Thierry Azaïs, and Gervaise Mosser1

From the Laboratoire de Chimie de la Matière Condensée de Paris, UMR 7574 Université Pierre et Marie Curie (UPMC)/Centre National de La Recherche Scientifique (CNRS)/Collège de France, UPMC, 4 place Jussieu, 75005 Paris, France

Received for publication, June 11, 2012, and in revised form, January 7, 2013 Published, JBC Papers in Press, January 22, 2013, DOI 10.1074/jbc.M112.390146

Background: Collagen fibrillogenesis is a fundamental process both in vivo and in vitro.

Results: Fibrillogenesis increases the heterogeneity of conformations of side chain amino acids, impacting 40% of imino acids.

Conclusion: A temperature-induced-attractive force component comes from significant amount of the collagen amino acids.

Significance: This work brings new perspectives to studies of collagen assembly and interactions with other matrix molecules.

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-$^1$C solid-state NMR spectroscopy. We interpret $^1$C chemical shift 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. Moreover, based on $^1$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 the self-assembling process and fibrillar structure of collagen.

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 supercoiled into a right-handed helix. These polypeptide chains, composed of $\sim$1040 amino acids each (1), are formed almost exclusively by repeats of three amino acid units, -(Gly-X-Y)$_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 these 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 and electron microscopy studies (5, 6). So far, the highest resolution obtained on the whole molecule by x-ray diffraction 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 Protein Data Bank), 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, reported $^1$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 (11–13). To deduce the backbone conformation, Saito et al. (9) have compared $^1$C chemical shifts of fibrillar collagen with those obtained on small triple helical peptides. Torchia et al. (14) have compared native and denatured resonances of collagen fragment CB2. More recently, Zhu et al. (10) have studied structural modifications of bovine cortical bone induced by dehydration. Even though all of these studies reveal important structural characteristics of collagen, many $^1$C chemical shifts were “hidden” due to a poor spectral resolution.

In the present work, using solid-state NMR, we have investigated structural changes in collagen at high concentrations in four different regimes: three native conditions defined as (i) soluble at acidic pH (pH 2.5; $T = 20 ^\circ$C), (ii) and (iii) fibrillar at basic pH (pH 8.5; $T = 20 ^\circ$C and $T = 30 ^\circ$C), and a denatured condition (iv) (pH 8.5; $T = 35 ^\circ$C). This last sample was used as a reference. We have observed spectrally resolved one-dimensional and two-dimensional $^1$H-$^1$C magic angle spinning (MAS)2 spectra ($^1$C typical line width from 20 to 40 Hz). We have identified side chains and backbone $^1$C chemical shifts modifications upon fibrillogenesis passing from acidic to basic.

* This work was supported by Ecole Doctorale Iviv 387 and the Université Pierre et Marie Curie (to P. D. S. P.)

** This article contains supplemental Tables S1–S4, Figs. S1–S7, equations, and data.

1 To whom correspondence should be addressed: Laboratoire de Chimie de la Matière Condensée de Paris, UMR7574 UPMC/CNRS/Collège de France, 75005 Paris, France. Tel.: 33-14-427-6553; Fax: 33-14-427-6539; E-mail: gmosser@snv.jussieu.fr.

2 The abbreviation used is: MAS, magic angle spinning.
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 described previously (6). Collagen concentration was assessed by hydroxyproline titration (16), and purity was determined by SDS-PAGE. Collagen stock solution (500 mm CH$_2$COOH, 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 $\times$ g at 10°C, variable angle centrifuge EPENDORF 5702R$^{\text{TM}}$). The pH was further increased by exposing the acidic sample to ammonia vapors (30%) at 20°C for 7 min. Two-dimensional solid-state NMR spectra were recorded 6 h after the sample exposition to ammonia and acquired during 6 h. The stability of the NMR spectra were recorded 6 h after the sample exposition to vapors (30%) at 20°C for 7 min. Two-dimensional solid-state NMR spectra were recorded 6 h after the sample exposition to ammonia and acquired during 6 h. The stability of the sample was checked by recording one-dimensional $^{13}$C spectra prior and after each two-dimensional $^1$H-$^{13}$C experiments. The spectra were similar for each condition. Finally, the “basic pH” sample was centrifuged-filtered to collect enough liquid ($\sim 150 \mu$l) 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 ($\nu^1(1H)$, 60 kHz) for 16 h.

**NMR Parameters—**Solid-state NMR analyses were performed on a Bruker Avance III 500 spectrometer at 11.7 Tesla operating at 500.21 MHz for $^1$H and 125.79 MHz for $^{13}$C. A Bruker MAS double resonance probe was used with 7-mm rotors spun at a frequency $v_{\text{MAS}} = 2.5$ kHz. A BCU-X unit was used to regulate the temperature between 20 and 35°C. For all experiments, $t_{\text{off}}(^1\text{H})$ and $t_{\text{off}}(^{13}\text{C})$ were 11 and 7 $\mu$s, respectively. Relaxation delays of 1 and 3 s for $^1\text{H}$ and $^{13}\text{C}$, respectively, were sufficiently long to record quantitative spectra. 512 scans were acquired for $^{13}$C single pulse experiment, and 48 scans were acquired for the two-dimensional $^1$H-$^{13}$C INSensitive Nuclei Enhanced by Polarization Transfert 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 $1/2v_{\text{CH}_4}$ value of 145 Hz and the refocusing delay (1.1 ms) was matched to get the CH, CH$_2$, and CH$_3$ resonances similarly phased. 256 $t_t$ transients were collected for two-dimensional $^1$H-$^{13}$C INEPT spectra. $^1$H and $^{13}$C chemical shifts were referenced (δ = 0 ppm) to tetramethylsilane. No $^{13}$C signal was detected through a cross-polarization MAS experiment (contact time of 10 ms, 440 scans, and a $^1$H decoupling of 40 kHz). See supplemental Fig. S2 for details.

Assignment and Fit—$^1$H and $^{13}$C chemical shift assignments were based on two-dimensional $^1$H-$^{13}$C INEPT MAS spectra (see Fig. 1 and supplemental Fig. S1), and previous work 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 two-dimensional $^1$H-$^{13}$C INEPT MAS spectra obtained in native conditions. In our assignments, we did not consider tryptophans and cysteines because collagen I from rat tendon do not present either (1). We were not able to assign any resonance to methionine, histidine, and tyrosine due to their very small amounts ($<0.6\%$ of the total amino acid content). We were not able to assign 3-hydroxyprolines ($\sim 5\%$ of total hydroxyprolines content (18)) because 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.

One-dimensional $^{13}$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 one-dimensional $^{13}$C spectrum (supplemental Fig. S2). However, we used Arg Cδ peaks, free of overlap in the one-dimensional spectra, to deduce the relative intraresidual intensity of Arg Cγ (supplemental Table S3 and supplemental data). Because there are two times less arginine 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 S.D. based on the match between experimental data and peak simulation. Supplemental Table S1 shows that S.D.s for these fits spread from 1 to 10% depending on carbon type and physicochemical condition. Noteworthy, in our case, a change of 0.01 to 0.02 ppm in peak position always induced an increase of the S.D. of $\sim 1$ 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 (S.D. 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. As 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 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 one-dimensional and one-dimensional $^1$H-$^{13}$C INEPT MAS spectra of a fibrillar sample are depicted in supplemental Fig. S2. These spectra present a remarkable spectral resolution (full width at half-maximum $\sim 0.2$ to 0.4 ppm) compared with 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 one-dimensional $^1$H-$^{13}$C INEPT spectrum are identical (supplemental Fig. S2). However, no signal was detected from a $^1$H cross-polarization MAS experiment (supplemental Fig. S2), 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 two-dimensional
Fibrillogenesis of Collagen I Studied by Solid-state NMR

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

\[ \text{\{1H\}-13C INEPT MAS experiment to correlate the } ^{1} \text{H} \text{ to the } ^{13} \text{C} \text{ chemical shifts (supplemental Fig. S3). } ^{1} \text{H} \text{ and } ^{13} \text{C} \text{ 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 and supplemental Fig. S1, bottom row). Backbone and side chain } ^{13} \text{C} \text{ chemical shifts found in the present work are in good agreement with data extracted from a random coiled protein library (S.D. 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 and supplemental Fig. S1, second row from bottom). Interestingly, increasing the pH induces chemical shift dispersion for a same carbon (Fig. 1 and supplemental Fig. S1, third row from bottom). For some carbons, the peak splitting is further increased by increasing temperature from 20 to 30 °C (Fig. 1 and supplemental Fig. S1, top row). In native conditions, most } ^{13} \text{C} \text{ are composed of more than one peak but, because of the poor } ^{13} \text{C} \text{ 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 } ^{13} \text{C} \text{ although we cannot quantify it. For all native conditions, Gly } ^{13} \text{C} \text{ displays two resonances centered at 43.37 ppm and 42.60 ppm with the same relative intensity and at quite close position compared with the denatured state (iv) (supplemental Figs. S1 and S4). In the proton dimension, the Gly } ^{1} \text{H} \alpha \text{ less intense peak presents a clear upfield shift } (\Delta \delta = −0.1 \text{ ppm}) \text{ compared with (iv) (supplemental Figs. S1 and S4). Considering the native conditions, the only } ^{13} \text{C} \text{ that presents a measurable change in chemical shift in the carbon dimension is Asp } ^{13} \text{C} \alpha, \text{ for which an increase in pH induces a clear downfield shift of } \Delta \delta = +0.15 \text{ ppm (supplemental Fig. S1). This shift difference is in agreement with that expected considering deprotonation effects (24, 25). Among } ^{13} \text{C} \text{ carbons, Asp } ^{13} \text{C} \beta \text{ shows an important downfield shift of } +0.6 \text{ ppm with the pH increase and, at 30 °C, pH 8.5 ± 0.5, it is composed of four peaks (Fig. 1C). Ala } ^{13} \text{C} \beta \text{ presents two peaks at acidic pH and increasing pH induces the apparition of one less intense peak at a downfield position (Fig. 1A and supplemental Table S3). For } ^{13} \text{C} \text{ possessing a hydroxyl group (Ser, Thr), no significant changes were seen.} \]

To better describe the effects of pH and temperature, we 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 and supplemental Fig. S1, second row from top). The second class contains only Pro and Hyp residues. The increase in pH and temperature induces an important (\( \Delta \delta \sim −0.6 \text{ ppm} \)) 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 and 30 °C (Fig. 1C). No significant changes are observed for Hyp Cγ (supplemental Fig. S1) 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 (Cβ 1 and 2) and Thr (Cγ 2) that present less intense peaks and show a slight increase in chemical shift dispersion (Fig. 1A). The effect of physicochemical conditions on the resonance dispersion of arginine Cγ, proline Cγ, and hydroxyproline Cβ is shown on supplemental Fig. S5. Furthermore, Fig. 2, A–C, compares quantitatively the increases in 13C 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 and supplemental Fig. S5, B and C). Because our one-dimensional 13C experiments were made under quantitative conditions, it is worthwhile to compare the signal intensities between the two-dimensional INEPT experiment (sensitive to local motion) and the one-dimensional quantitative experiment. The difference between two-dimensional and one-dimensional quantitative experiment is relatively small (<20%) (supplemental Fig. S2 and Table S3).

DISCUSSION

Collagen Dynamics—The spectra obtained through two-dimensional (1H)-13C INEPT MAS and one-dimensional 13C experiments are in good agreement with previous 13C NMR experiments reported on native and denatured samples (14). We observe that varying the pH does not impact significantly the 1H 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 1H resonances (supplemental Table S1). Mosser et al. (26) 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 (~5–15 nm of diameter). Nano-fibrils are also present within larger collagen I fibrils (5). As in this study we are dealing with very small fibrils submitted to fast dynamics, this explains why the spectral resolution and the INEPT signal intensity are the same in the soluble and the fibrillar state. 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 nano-fibrils must be more constrained (for example, 50 to 200 nm in Ref. 27). This also explains why we were unable to record any 13C signal through a cross-polarization 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 two-dimensional INEPT spectra is always close to that found in one-dimensional quantitative spectra (supplemental Table S3). 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 (χ1, χ2, χ3) 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 13C chemical shift (supplemental Fig. S6) (15). For proteins, this effect overcomes all other effects such as the ring currents (29). Because at room temperature amino acid side chains of rat tendon fibrils show a very fast interconversion rate between different conformations (11–13), each 13C chemical shift in our spectra must represent an equilibrium between those conformations in specific locations along the polypeptide chain. In the denatured condition (iv), when no specific protein fold is present, almost all carbons display only one 13C 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 study, which deals with the full collagen triple helix, Arg Cγ 13C 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 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 one-dimensional quantitative experiment, all of 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 six major helical conformations in the human collagen I triple helix. This work shows that for regions poorer in imino acids, the triple helical twist is less important (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. 1A and supplemental Table S2). The Ala Cβ peak at 16.6 ppm is not affected by the association into triple helix because this chemical shift remains comparable with that of (iv) (Fig. 1A and supplemental Table S2). In random coiled peptides, residues immediately followed by a proline present a chemical upfield shift of ~1 ppm for Cα and Cβ carbons (23). Because the peak fitting for (iv) (supplemental Table S2) agrees well with the number of Ala followed by an imino acid 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 other peaks, 17.14, 17.52, 17.85 ppm, correspond to regions rather depleted in imino acids and the 18.2-ppm peaks correspond to the richest region.

Side Chain Analysis (Equilibrium Quantification)—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 Protein Data Bank with 13C 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 these data from the library, 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 χ1 adopts the trans (180°) or −gauche (60°) conformation, the Gln Cγ chemical shift is ~0.5 ppm lower than when χ1 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 χ1. 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 Cδ1 conformers. These calculated chemical shift differences are always close to the difference found between the two more extreme values of the London et al. 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 14C 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 physicochemical conditions are not the same (14, 23). We thus used 13C 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
TABLE 1

|           | Asparagine | Aspartic Acid | Glutamine | Glutamic Acid |
|-----------|------------|---------------|-----------|---------------|
| pH 6.5    | 37.0       | 39.2          | 39.7      | 40.5          |
| pH 8.5    | 37.1       | 39.3          | 39.8      | 40.6          |
| pH 8.5    | 37.2       | 39.4          | 40.0      | 41.0          |
| pH 7.5    | 37.3       | 39.5          | 40.1      | 41.1          |

The table shows the chemical shift (in ppm) for the amide proton of Asparagine, Aspartic Acid, Glutamine, and Glutamic Acid at different pH values. The chemical shift values are measured relative to TMS. The higher the chemical shift value, the greater the upfield shift from TMS.

The data indicates that the chemical shift values for Asparagine, Aspartic Acid, Glutamine, and Glutamic Acid increase with increasing pH. This is consistent with the expected behavior of amide protons, which become more protonated and thus experience a greater upfield shift with increasing pH.

Impact of pH on the Chemical Shift of Amide Protons

The chemical shift values for the amide proton of Asparagine, Aspartic Acid, Glutamine, and Glutamic Acid increase with increasing pH. This is consistent with the expected behavior of amide protons, which become more protonated and thus experience a greater upfield shift with increasing pH.

Amide protons are sensitive to the pH of the solution. As the pH increases, the amide protons become more protonated, resulting in a greater upfield shift. This effect is known as the amide shift effect.

The effect of pH on the chemical shift of amide protons is shown in the table above. The chemical shift values are measured relative to TMS. The higher the chemical shift value, the greater the upfield shift from TMS.

The data indicates that the chemical shift values for Asparagine, Aspartic Acid, Glutamine, and Glutamic Acid increase with increasing pH. This is consistent with the expected behavior of amide protons, which become more protonated and thus experience a greater upfield shift with increasing pH.

The chemical shift values for Asparagine, Aspartic Acid, Glutamine, and Glutamic Acid increase with increasing pH. This is consistent with the expected behavior of amide protons, which become more protonated and thus experience a greater upfield shift with increasing pH.

The chemical shift values for Asparagine, Aspartic Acid, Glutamine, and Glutamic Acid increase with increasing pH. This is consistent with the expected behavior of amide protons, which become more protonated and thus experience a greater upfield shift with increasing pH.

Impact of pH on the Chemical Shift of Amide Protons

The chemical shift values for Asparagine, Aspartic Acid, Glutamine, and Glutamic Acid increase with increasing pH. This is consistent with the expected behavior of amide protons, which become more protonated and thus experience a greater upfield shift with increasing pH.

The chemical shift values for Asparagine, Aspartic Acid, Glutamine, and Glutamic Acid increase with increasing pH. This is consistent with the expected behavior of amide protons, which become more protonated and thus experience a greater upfield shift with increasing pH.

The chemical shift values for Asparagine, Aspartic Acid, Glutamine, and Glutamic Acid increase with increasing pH. This is consistent with the expected behavior of amide protons, which become more protonated and thus experience a greater upfield shift with increasing pH.
helices. However, our data show that these numerous combinations of neighboring amino acids in the quaternary structure induce only a restricted number of side chain conformations. The present data also suggest that the relatively weak temperature-induced attractive force component responsible for fibrillogenesis comes from a rather large proportion of amino acids of the molecule.

Actually, previous force measurement studies, at neutral pH, demonstrated that the self-assembling net force of collagen 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. Because 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 <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 have a similar chemical shift to those found in the denatured state i.e. about one-half 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 chain/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 that show an increase in equilibrium dispersion with both temperature and pH, must necessarily be located in regions where energetically favored complementary surfaces are formed. Moreover, considering all side chains that answer this criteria (equilibrium changes with T and pH), one can estimate that these correspond to 10 to 15% of the total collagen residue content (supplemental Table S1). 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.

CONCLUSIONS

$^{13}$C and $^{1}$H-$^{13}$C solid-state MAS NMR experiments were used to identify conformational changes in collagen I in three 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, with 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 the literature, we assigned changes in $^{13}$C chemical shift to the most spread-out side chain conformations and thus calculated equilibrium distributions between conformations. We also show that fibrils formation impacts on many imino acids, stabilizing prolines in the down pucker conformation and hydroxyproline both in up and down conformations. 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 thus 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 extracellular matrix molecules.

Acknowledgments—We thank Olivier Lequin for fruitful discussions at an early stage of manuscript preparation and the reviewers and Sandrine Quignard for reading and revising the manuscript.

REFERENCES

1. Bolboaca, S. D., and Jäntschl, L. (2007) Amino acids sequence analysis on collagen. Bulletin Uso Samm-on 63–64, 311–316
2. Berisio, R., De Simone, A., Ruggiero, A., Impura, R., and Vitagliano, L. (2009) Role of side chains in collagen triple helix stabilization and partner recognition. J. Pept. Sci. 15, 131–140
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., and 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., Irving, T. C., Miller, A., and Wess, T. 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. M., and 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. H., Fan, P., Brodsky, B., and 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. Saitō, H., Tabeta, R., Shoji, A., Ozaki, T., Ando, L., and Miyata, T. (1984) A high-resolution 13C-NMR study of collagenlike 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., and Ramamoorthy, A. (2009) Time-resolved dehydration-induced structural changes in an 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., and Torchia, D. A. (1982) Characterization of leucine side-chain reorientation in collagen-fibrils by
solid-state $^2$H 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., and 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., Jr., and Quattrone, A. J. (1975) Molecular dynamics and structure of the random coil and helical states of the collagen peptide, 03β11-CB2, as determined by 13C magnetic resonance. *Biochemistry* **14**, 887–900

15. London, R. E., Wingad, B. D., and 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., and Loxley, R. (1963) Two 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., Hudson, D. M., Wu, J. J., and 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) $^{13}$C n.m.r. chemical shifts in aliphatic alcohols and the γ-shift caused by hydroxyl and methyl groups. *Org. Magn. Reson.* **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., and Hoatson, G. (2002) Modelling one- and two-dimensional solid-state NMR spectra. *Magn. Reson. Chem.* **40**, 70–76

21. Reichert, D., Pascui, O., deAzevedo, E. R., Bonagamba, T. J., Arnold, K., and 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., and 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., and 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., and Oldfield, E. (1994) Chemical shifts of carbonyl carbons in peptides and proteins. *J. Am. Chem. Soc.* **116**, 11485–11488

25. Kohda, D., Sawada, T., and Inagaki, F. (1991) Characterization of pH titration shifts for all the nonlabile proton resonances 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., and 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., and 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., Jr., and Karplus, M. (1994) Conformational analysis of the backbone-dependent rotamer preferences of proteins side chains. *Nat. Struct. Biol.* **1**, 334–340

31. Mikhailov, D., Daragan, V. A., and 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, S. D., Kollman, P. A., and Klein, T. E. (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., and Nishino, N. (2004) Crystal structures of collagen model peptides with Pro-Hyp-Gly repeating sequence at 1.26 Å resolution: implications for proline ring puckering. *Biopolymers* **76**, 367–377

34. Leikin, S., Rau, D. C., and Parsegian, V. A. (1994) Direct measurement of forces between self-assembled proteins: temperature-dependent exponential forces between collagen triple helices. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 276–280

35. Leikin, S., Rau, D. C., and Parsegian, V. A. (1995) Temperature-favoured assembly of collagen is driven by hydrophilic not hydrophobic interactions. *Nat. Struct. Biol.* **2**, 205–210