Trimeric Assembly and Three-Dimensional Structure Model of the FACIT Collagen COL1−NC1 Junction from CD and NMR Analysis†

Anne Lesage,‡ François Penin,‡ Christophe Geourjon,‡ Dominique Marion,§ and Michel van der Rest*,§

Institut de Biologie et de Chimie des Protéines, CNRS UPR 412, 7 passage du Vercors, 69367 Lyon Cedex 07, France, and Institut de Biologie Structurale Jean-Pierre Ebel, CNRS-CEA, 41 avenue des Martyrs, 38027 Grenoble Cedex 1, France

Received November 8, 1995; Revised Manuscript Received May 14, 1996⊗

ABSTRACT: The 3D structure of the COL1−NC1 junction of FACIT type XIV collagen was investigated using GYCDPSSCAG and $(GPP^*)_3$GYCDPSSCAG synthetic peptides, circular dichroism, and NMR. At −20 °C and under air oxidation catalyzed by Cu^{2+}, the peptide $(GPP^*)_3$GYCDPSSCAG is able to self-associate with high yield into a stable triple disulfide bonded trimer. The presence of a triple helical conformation was confirmed by circular dichroism. The analysis of the trimer by 2D NMR provided a set of distance constraints for the noncollagenous part. Molecular models for the 3D structure of COL1−NC1 junction were calculated, using the NMR distance constraints in combination with the 3D structural data recently established by X-ray crystallography [Bella, J., Eaton, M., Brodsky, B., & Berman, H. M. (1994) *Science 266*, 75−81] for a collagenous triple helix. From the eight theoretically possible arrangements for the three interchain disulfide bonds, only two close disulfide conformers are compatible with the experimental data. The main feature of the trimer structure is the asymmetry of the molecule due to the disulfide bond pattern that induces a particular folding of one chain. This chain forms a turn-like structure locked by two disulfide bonds with the two other chains. The turn-like folding is close to that observed for the cyclized oxidized monomeric peptide. This is the first report of the 3D structure model for a junction between a collagenous triple helical domain and a noncollagenous domain.

Collagens are proteins of the extracellular matrix which include one or more domains that have a triple helix conformation with a repetitive $(Gly\text{-}X\text{-}Y)_n$ primary sequence (van der Rest & Garonne, 1991). About 30% of positions X and Y are occupied by prolyl and hydroxyprolyl residues, respectively, which stabilize the helical conformation. Nineteen collagens have been defined so far (Mayne & Brewton, 1993). The FACITs (fibril-associated collagens with interrupted triple helix) which include type IX, XII, XIV, XVI, and XIX are characterized by short triple helical domains (named COL domains) separated by non triple helical domains (named NC domains). Type XVI and XIX collagens have only been characterized at the nucleic acid and genomic levels, and their molecular organization has not been established (Pan et al., 1992; Myers et al., 1994). They will not be considered further in this study. Type IX collagen is a heterotrimer α1(IX)α2(IX)α3(IX), while type XII and XIV collagens are homomeric molecules. The C-terminal end of the FACITs is composed of a noncollagenous domain NC1 preceded by a highly conserved triple helical domain COL1. All the FACITs present a remarkable sequence homology at the junction COL1−NC1. Two cysteine residues separated by four amino acids are strictly conserved (Ninomiya & Olsen, 1984; Ninomiya et al., 1985; Gordon et al.,1989; Kimura et al., 1989; Muragaki et al., 1990; Dublet & van der Rest, 1991; Gordon et al., 1991; Brewton et al., 1992; Abe et al., 1994). The first cysteine is part of the last tripeptide of the COL1 domain at position Y while the second one is located in the NC1 domain. Between the two, there are the following: in position 1, a charged residue, generally an acidic one; in position 2, most often a proline; in positions 3 and 4, small amino acids (Ser or Ala). These cysteine residues are known to be involved in interchain disulfide bonds (Dublet & van der Rest, 1987, 1991; van der Rest & Mayne, 1987), but there are no data available concerning the disulfide bonding pattern(s).

The mechanisms of selection, correct registration, and assembly of the three polypeptidic chains remain unknown for the FACITs. Several recent results of our group suggest that the COL1 domain and the cysteine-containing region may be involved in those mechanisms (Labourdette & van der Rest, 1993; Mazzorana et al., 1993). To understand the role of the COL1−NC1 junction in chain assembly, the knowledge of its three-dimensional structure is required. Only very few studies were previously published concerning the arrangement of interchain disulfide bonds in collagenous molecules. A theoretical space-filling model was proposed for the junction between the long triple helix and the C-terminal noncollagenous domain of type III collagen (Bruckner et al., 1978). In this collagen, the three interchain disulfide bonds are formed between adjacent cysteine residues. However, in the case of FACIT molecules, where the two cysteines are separated by four amino acids in each chain, a theoretical molecular modeling is not trivial. Taking into account the shift of one amino acid between the three chains, necessary for the formation of the collagenous triple helix in the COL1 domain, there are eight ways of pairing the cysteines for three identical chains. For three distinct chains, no less than 48 arrangements must be considered.

† This work was supported by the Centre National de la Recherche Scientifique.

* To whom correspondence should be addressed. E-mail: mv@ibcp.fr. FAX: (33) 72-72-26-02.

‡ Institut de Biologie et de Chimie des Protéines.

§ Institut de Biologie Structurale.

⊗ Abstract published in *Advance ACS Abstracts,* July 1, 1996.

S0006-2960(95)02666-3 CCC: $12.00

Thus, experimental structural data were needed to guide the molecular modeling of the COL1−NC1 junction.

As FACITs are minor collagens in the matrix, the preparation of the COL1−NC1 junction from tissues in sufficient amounts for structural studies would be unpractical. To overcome this problem, we synthesized the two peptides, GYCDPSSCAG and $(GPP^*)_3$GYCDPSSCAG: the GYCDPSSCAG sequence corresponds to the COL1−NC1 junction of chicken α1(XIV) collagen and for the longer peptide, the three GPP* repeats mimic the COL1 domain. Although the natural sequence of the C-terminal region of the COL1 domain of collagen XIV is GSQGPAGPP*, we have chosen the repetitive sequence GPP*GPP*GPP* for the following reasons. (1) This region is triple helical in the native protein as shown by its resistance to proteolytic cleavage (Dublet & van der Rest, 1991), but it is part of a larger triple helical sequence that contributes to the stability of the helical conformation. The chosen sequence would be more favorable for a triple helical folding of a short peptide. (2) The GPP*GPP*GPP* sequence would give simple NMR[1] spectra in the region of amide proton correlations since P and P* do not have an amide proton. (3) The GPP*GPP*GPP* sequence naturally occurs at the end of the COL1 domain in other FACIT collagens such as the α1(IX) chain (Ninomiya & Olsen, 1984). (4) The three-dimensional structure of a GPP*GPP*GPP* sequence in its triple helical conformation has been previously described (Bella et al., 1994). We report here the ability of the peptide $(GPP^*)_3$GYCDPSSCAG to self-associate specifically into a disulfide-bonded trimer as identified by mass spectroscopy. Characterization by microsequencing, circular dichroism, and 2D NMR shows that this trimer exhibits a specific asymmetric conformation. NMR spectral analysis allows us to obtain sequence-specific assignments as well as a set of distance constraints in the noncollagenous part. These constraints were used in combination with the recent 3D structural data of the collagenous triple helix established from X-ray crystal diffraction (Bella et al., 1994) to simulate the structure of the COL1−NC1 junction. To our knowledge, this study is the first report of a 3D structural analysis for a disulfide-linked junction between a collagenous triple helical domain and a noncollagenous one.

MATERIALS AND METHODS

Peptide Synthesis. Peptide synthesis was done according to standard protocols using 9-fluorenylmethoxycarbonyl (Fmoc) chemistry on a Fmoc-L-Gly-KA resin in a Milligen 9050 pepsynthesizer. The triplets GPP*, purchased from Bachem Bioscience, were added in a single step using the Fmoc-Gly-Pro-Hyp-OH tripeptide (Bernillon and Wallach, personal communication).

Concentration Measurements. Peptide concentrations in solution at neutral pH were estimated from UV absorbance of tyrosine using the molar extinction coefficient at 280 nm of 1536 M^{-1} cm^{-1} (Beaven & Holidays, 1952).

Thiol Titration. The extent of cysteine oxidation was monitored by RP-HPLC after incubating the peptide (1 mg/mL) in the presence of 10 mM *N*-ethylmaleimide (NEM) for 15 min at room temperature (Lee et al., 1992). The titration of thiol groups in the trimeric assembly was done according to Ellman's method (Riddles et al., 1983) by incubating the disulfide-linked trimer in a freshly prepared solution of DTNB [5,5′-dithiobis (2-nitrobenzoic acid)] and measuring the net absorbance at 412 nm after 5 min of incubation at room temperature (molar extinction coefficient of 13 600 M^{-1} cm^{-1}).

Microsequencing. The location of disulfide bonds was done by microsequencing using an Applied Biosystems 470A sequencer, as previously reported (Marti et al., 1987). The Edman degradation of cystine gives two compounds corresponding to the di-PTH-cystine and its derivative dehydroalanine. Insulin was used to calibrate the microsequencing yields and the retention times of di-PTH-cystine and dehydroalanine.

Circular Dichroism. CD spectra were recorded on a CD6 Jobin Yvon spectropolarimeter equipped with a variable temperature unit. Measurements were done with a 1 mm path-length cuvette at a constant rate of 1 nm/min with a 0.2 nm resolution. Thermal transition curves were obtained by monitoring $[\theta]_{228nm}$ as a function of temperature.

NMR Spectroscopy. All NMR experiments were performed on a Varian UNITY*plus* 500 spectrometer, equipped with *ultra-nmr* shims and using a pfg proton−carbon−nitrogen 5 mm probe. Phase-sensitive DQF-COSY (Rance et al., 1983), TOCSY (Braunschweiler & Ernst, 1983), NOESY (Jeener et al., 1979; Kumar et al., 1980), and ROESY (Bax & Davis, 1985a) experiments were performed accordingly. For NOESY experiments, two mixing times of 150 and 250 ms were used. TOCSY experiments were carried out using the MLEV-17 isotropic transfer sequence (Bax & Davis, 1985b) and a mixing time of 80 ms. A continuous irradiation *spin-lock* was used for ROESY experiments. In all 2D experiments, quadrature detection in the indirectly observed dimension was obtained with States−TPPI (Marion et al., 1989). Solvent suppression was carried out using selective, low-power irradiation during the 1.5 s relaxation delay and, for NOESY experiments, during the mixing time. The spectral width used in both dimensions was 5000 Hz; 2K complex data points in t_2 and 512 experiments in t_1 (zero-filled to 1K) were acquired. The data sets were processed using VNMR software. Alternatively, NMR spectra of the trimer were also treated with GIFA software utilities (M. A. Delsuc, Oxford Molecular, version 3.01), which allow 2D inverse Fourier transforms. Squared sine-bell apodization functions were used in both dimensions, shifted to closely match the shape of the FIDs. Baseline corrections were performed using a spline fitting of predefined baseline regions. The assignment of ^{13}C resonances was based on gradient-HSQC experiments (Ruiz-Cabello et al., 1992). The strength of the two gradients, inserted at the end of the evolution time t_1 and after the 180° pulse of the

[1] Abbreviations: ACN, acetonitrile; CD, circular dichroism; DMSO, dimethyl sulfoxide; DQF-COSY, double-quantum-filtered correlation spectroscopy; DTT, dithiothreitol; FID, free induction decay; GSH and GSSG, the reduced and oxidized forms of glutathione, respectively; HPLC, high-performance liquid chromatography; RP-HPLC, reversed-phase high-performance liquid chromatography; HSQC, heteronuclear single-quantum coherence; gHSQC, gradient heteronuclear single-quantum coherence; NEM, *N*-ethylmaleimide; NMR, nuclear magnetic resonance; ppm, parts per million; 2D, two dimensional; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; pfg, pulse field gradient; PTH, phenylthiohydantoin; RMSD, root mean square deviation; ROESY, rotating frame Overhauser effect spectroscopy; TFA, trifluoroacetic acid; TMS, tetramethylsilane; TOCSY, total correlation spectroscopy. Letters P and P* are used for proline and hydroxyproline, respectively; Hyp is hydroxyproline; Xn_N corresponds to the amino acid X in position *n* of peptidic chain *N*.

retro-inept block, was 28 G/cm, and their respective durations were 2 and 0.5 ms. Quadrature detection was achieved by cycling the sign of the second gradient. The spectral widths were 18 and 5 kHz in the ^{13}C and ^{1}H dimensions, respectively. The spectra were collected with 2K data points in the t_2 dimension and 256 t_1 increments; 128 scans were acquired for each t_1 value. Proton and carbon chemical shifts were referenced to TMS.

Computer Modeling. For the molecular modeling of the oxidized peptide GYCDPSSCAG, 17 distance constraints were derived from NOE intensities (mixing time of 150 ms) and classified into three categories: <2.8, <3.6, and <5 Å. An additional distance constraint of 2.02 Å between the two sulfur atoms was introduced. To model the triple helical part of the trimeric peptide (residues 1−9 of each chain), the coordinates of the three N-terminal GPP* triplets of the collagen-like peptide $(PP^*G)_4PP^*A(PP^*G)_5$ were used to calculate interproton distances (Bella et al., 1994; Protein Data Bank code number 1CAG). Only distances ≤3 Å and involving either an amide or/and an α proton were considered and used as distance constraints in X-PLOR software (see below). Each value was allowed to vary from −0.3 to +0.1 Å around the actual interproton distance. In this way, a total of 39 constraints were introduced for the triple helical region, 20 of which being interchain constraints. In addition, average (Φ, Ψ) dihedral angles deduced from the crystal structure were also introduced with the following values: for Gly, $\Phi = -71 \pm 20°$ and $\Psi = 174 \pm 25°$; for Pro, $\Phi = -72 \pm 15°$ and $\Psi = 164 \pm 20°$; for Hyp, $\Phi = -60 \pm 15°$ and $\Psi = 150 \pm 18°$ [Table 2 in Bella et al. (1994); note that the reported standard deviations were doubled for each angle]. For the C-terminal part of the molecule (residues 10−19 of each chain), distance constraints were derived from the NOESY spectra with a mixing time of 150 ms.

Three-dimensional structures were generated from NOE distances with the X-PLOR 3.1 program (Brünger, 1992) using the default parameter sets, except for some minor modifications to increase the durations of the molecular dynamic simulations and the number of energy minimization steps. Briefly, a template structure with randomized coordinates and ideal geometry was generated (protocol "generate_template.inp", using the all-hydrogen force fields "topallhdg" and "parallhdg"). Then a substructure distance geometry protocol was applied ("dg_sub_embed.inp"), introducing the NOE-derived distance constraints to generate 100 unregularized structures with a partial set of coordinates ("embedded substructures"; protons without stereospecific assignment were treated as pseudoatoms, and the correction factors were added to the upper and lower distance constraints; Wüthrich et al., 1983). The coordinates of these substructures were then regularized by simulated annealing (Nilges et al., 1988a,b) using the "dgsa.inp" protocol, consisting of 9 ps of heat procedure up to 2000 K by increasing steps of 50 K. The contribution of the van der Waals terms was reduced during this stage to allow a maximal exploration of the conformational space. The system was cooled down to 250 K by steps of 25 K in 10 ps with a concomitant increase of the van der Waals interactions. In a following step, the structures were refined by the application of the protocol "refine.inp", consisting of simulated annealing and final energy minimization. In this step, the duration of the high temperature stage was 50 ps, and the temperature was decreased from 1000 K down to 250 K by steps of 25 K. Each structure was then submitted to 5000 cycles of Powell energy minimization. A final minimization step was applied with the standard force field parameters except that K_{bonds} and K_{angles} were fixed to 600 kcal·mol^{-1}·Å^{-2} and 90 kcal·mol^{-1}·rad^{-2}, respectively. When the position of the three disulfide bonds was fixed, additional distance constraints of 2.02 Å between the pairs of sulfur atoms were introduced. The disulfide bonds were considered as covalent bonds only during the refinement protocol (refine.inp) and final minimization (minimize.inp).

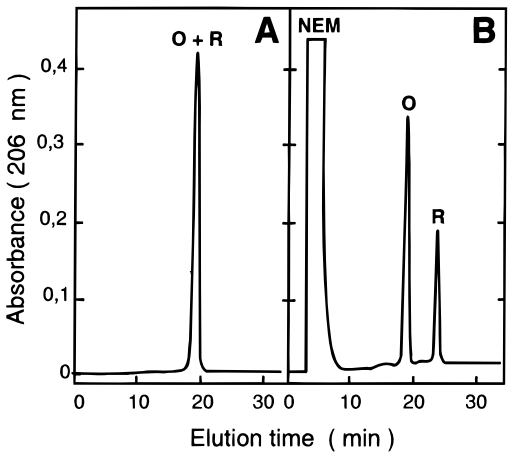


FIGURE 1: Separation of the reduced (R) and oxidized (O) forms of the peptide $(GPP^*)_3$GYCDPSSCAG after modification of the reduced peptide by NEM. The peptide $(GPP^*)_3$GYCDPSSCAG was incubated during 24 h at a concentration of 1 mg/mL at room temperature in 50 mM Tris-HCl and 100 mM NaCl, pH 8.5, buffer. The solution was analyzed on an ABI C18 RP-HPLC column (220 × 2.1 mm) using a 0.5%/min linear gradient of acetonitrile in 0.1% TFA at a 0.25 mL/min flow rate. (A) The oxidized and reduced monomeric forms are coeluted (retention time at 19 min). (B) Prior to injection onto the column, a 20 molar excess of NEM/mol of cysteine residue was added to the solution. The reduced peptide covalently modified by NEM is eluted latter off the column (retention time 24 min) than the unmodified oxidized one (retention time 19 min).

Ensembles of 100 structures were calculated in order to sample as widely as possible the conformational space, and the structures were compared on the basis of pairwise RMSDs for the backbone atom coordinates (N, Cα, and C′). Local analogies were analyzed by considering the local RMSD plot (Berndt et al., 1993). By sliding a tripeptide window along the sequence, variations in the structures were detected and analyzed. Structure superimposition, 3D graphic displays, and manipulations were accomplished using ANTHEPROT 2.0 software (Geourjon & Deléage, 1995).

RESULTS

Peptide Characterization. The synthetic peptides GYCDPSSCAG and $(GPP^*)_3$GYCDPSSCAG corresponding to the COL1−NC1 junction of chicken type XIV collagen were purified by RP-HPLC (elution at 15% and 21% ACN, respectively). The extent of cysteine oxidation by air into an internal disulfide bond was measured for both peptides after *N*-ethylmaleimide (NEM) derivatization. Indeed, as illustrated in Figure 1 for the peptide $(GPP^*)_3$GYCDPSSCAG, while the oxidized and reduced monomeric forms of the peptide coelute (Figure 1A), the NEM-modified peptide corresponding to the reduced form is eluted later (Figure 1B). Such analysis showed that (i) the peptides purified from the crude product of synthesis were totally in a reduced form

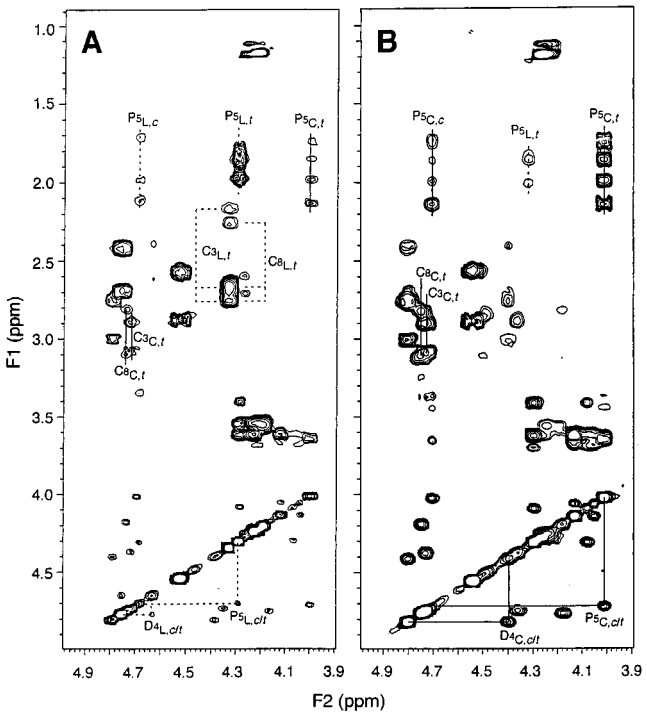


FIGURE 2: Contour plots of phase-sensitive 500 MHz TOCSY experiments of the peptide GYCDPSSCAG (A) collected as soon as possible from a fresh preparation of the reduced peptide and (B) at equilibrium. The concentration was 1.5 mM in DMSO-d_6 and the temperature 30 °C. A 2 ms trim pulse preceded the 80 ms MLEV-17 spin-lock. The exchange cross-peaks of $D4_{H\alpha}$ and $P5_{H\alpha}$ due to *cis*−*trans* proline isomerization as well as the resonances of side chain protons for P5, C3, and C8 are indicated. Dotted and solid lines are used for the linear and cyclic conformations respectively. L = linear conformation; C = cyclic conformation; suffix *t* or *c* refers to the *trans* or *cis* isomery of the proline residue.

and (ii) the extent of peptide air oxidation investigated in several solvents occurred at a relatively low rate (half-time of several hours), indicating that the reactivity of cysteines residues is moderate for both peptides.

The reduced and oxidized forms of the peptides were both characterized by ^{1}H 2D NMR spectroscopy. From freshly prepared samples of the reduced peptides, NMR TOCSY, DQF-COSY, NOESYm and ROESY spectra, collected at various times over 70 h, showed that the peptides adopt several structures due to cysteine oxidation and proline *cis*−*trans* isomerisation. As illustrated in Figure 2 for the peptide GYCDPSSCAG on TOCSY spectra, we observed the time disappearance of a first set of resonances to the benefit of a second one. This corresponds to the conversion of the reduced, linear form (L) to the oxidized, cyclized one (C). Indeed, evidence for peptide cyclization is (i) the disappearance of thiol proton resonances (observed at 2.18 and at 2.27 ppm for $C3_{L,t}$ and $C8_{L,t}$, respectively in Figure 2A but absent for $C3_{C,t}$ and $C8_{C,t}$ in Figure 2B) and (ii) the observation of several medium-range NOE correlations for the second set of resonances between residues D4 and S6−S7 (Figure 3), indicating the formation of a cycle between the two cysteines.

TOCSY spectra also showed exchange cross-signals between Hα resonances. They are indicated for residues D4 and P5 in Figure 2. The cross-signals corresponding to the linear form (Figure 2A) disappeared over 30 h while those corresponding to the cyclized form became more intense (Figure 2). These cross-signals are due to chemical exchange as checked on ROESY spectra where they have the same

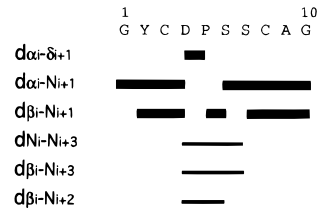


FIGURE 3: Summary of the backbone NOE connectivities observed for the oxidized GYCDPSSCAG peptide having a *trans* proline. NOE intensities were measured as described in Materials and Methods in NOESY spectra (mixing times of 150 and 250 ms, 30 °C in DMSO-d_6). The thickness of the lines indicates the strong and weak NOE intensities.

sign as the diagonal (data not shown) and correspond to the *cis*−*trans* isomerization of P5. Sequential NOEs between $D4_{H\alpha}$ and $P5_{H\delta\delta'}$ were observed for both linear and cyclic major conformations, indicating that P5 adopts preferentially a *trans* configuration ($P5_{L,t}$ and $P5_{C,t}$). Both linear and cyclic minor conformations with a *cis* proline were visible ($P5_{L,c}$ and $P5_{C,c}$, Figure 2). Additional indications for *cis* isomers were the chemical shift of $P5_{H\alpha}$ which moves from 4.31 to 4.69 ppm for the linear conformer and from 4.02 to 4.72 ppm for the cyclic one. Indeed, the resonance of $Pro_{H\alpha}$ is shifted downfield in DMSO for a *cis* conformation on model peptides (Deber et al., 1970). Moreover, for cyclized peptide, NOESY and ROESY spectra showed exchange-mediated NOE correlations between $D4_{H\alpha}$ of the major *trans* isomer and $P5_{H\alpha}$ of the minor *cis* one as well as between $P5_{H\alpha}$ of the major one and $D4_{H\alpha}$ of the minor one. For the cyclic conformer, the absence of chemical exchange cross-signals for the N- and C-terminal amino acids G1, Y2, and G10 indicates that their environment is not affected by *cis*−*trans* isomerization. A table containing the proton chemical shifts of the assigned residues for the various structures of the peptide GYCDPSSCAG (linear and cyclic, with *cis* or *trans* proline) is available as Supporting Information.

Figure 3 summarizes the NOEs observed for the cyclized monomer with the *trans* isomery of P5. A set of 17 distance constraints deduced from the intensity of these NOE correlations as well as an additional constraint between the two sulfur atoms, introduced to set the disulfide bond, allowed the molecular modeling of the oxidized peptide. Figure 11A shows a stereoview of these structures after superposition. A RMSD of 0.22 Å was calculated on the peptidic backbone of 20 generated structures (residues 3−8). The negative χ_3 angles define a left-handed disulfide bond (Thornton, 1981). Similar structural models were obtained without introducing a distance constraint between sulfur atoms.

The characterization of the linear and cyclized forms of the monomeric peptide $(GPP^*)_3$GYCDPSSCAG was also done by NMR, although the presence of the three triplets GPP* and the *cis*−*trans* isomerization of its proline and hydroxyproline residues complicated the spectra and made especially difficult the assignment of Pro 14 (data not shown).

Trimeric Peptide Assembly. The above results show that the natural tendency of the monomers is to cyclize by cysteine oxidation. This constituted the main problem encountered for trimeric peptide assembly, which is stabilized by three disulfide bond formations. All attempts performed with the peptide GYCDPSSCAG to get specific formation of trimer with a good yield failed. More successful results were obtained with the peptide $(GPP^*)_3$GYCDPSSCAG. Several different approaches were tested to perform trimer

Table 1: Percentages of Association Products of $(GPP^*)_3GYCDPSSCAG$ Peptide Obtained for Various Incubation Temperatures[a]

	temperature of incubation (°C)			
	25	15	5	−20
M (%)	99.2	92.7	90.2	83.9
D (%)	nd	6.2	6.5	6.1
T (%)	nd	1.1	3.3	10

[a] The peptide $(GPP^*)_3GYCDPSSCAG$ was incubated for 15 h in the association buffer, 50 mM Tris-HCl, 100 mM NaCl, pH = 8.5, and 30% methanol, at a concentration of 1 mg/mL. At 25 °C, the percentages of trimer and dimer were not determined (nd) since they could not have been clearly distinguished. The percentages of the various association products were evaluated from RP-HPLC chromatograms. M = monomer; D = dimer; T = trimer.

assembly via disulfide bond formation, either by thiol−disulfide interchange reactions or by air oxidation of cysteine residues (Creighton, 1984).

In the first approach, oxidation attempts were made by incubating the peptide with a mixture of oxidized (GSSG) and reduced (GSH) glutathione as already described (Wetlaufer, 1984). Although some products corresponding to dimeric and trimeric peptides could be detected, it yielded a very low amount of associaton species (<5%) and was thus not useful to prepare large amounts of trimer (data not shown). As shown above, the air oxidation approach of $(GPP^*)_3GYCDPSSCAG$ peptide at 25 °C mainly leads to cyclization by formation of an intrachain disulfide bond. However, when the temperature of incubation is decreased, dimeric and trimeric species are formed. Table 1 shows the percentage of the different species obtained at various temperatures after 15 h of incubation. Peptide species in the mixture were separated by RP-HPLC (Figure 4A) and identified by electrospray mass spectroscopy (measured average molecular mass of 3515 ± 0.7 and 5274 ± 1.1 g/mol for species D and T of Figure 4A, respectively; calculated isotopic mass of 1759 g/mol for the reduced monomeric peptide). The formation of trimer increases when the temperature decreases while the dimer fraction remains sensibly constant (Table 1). At −20 °C, the formation of trimers becomes superior to that of dimers. As all experiments were done in the presence of 30% methanol added in the association buffer as an antifreeze, the formation of trimer is essentially due to the decrease of incubation temperature. Similar results were obtained with other alcohols (ethanol, 2-propanol) with an alcohol concentration above 30% and at temperatures below −20 °C. Thus, the decrease of the incubation temperature seems to be the only factor that increases the trimer formation. The incubation temperature was set to −20 °C in the following experiments.

The assembly kinetics of the $(GPP^*)_3GYCDPSSCAG$ peptide at −20 °C is complex as shown in Figure 4B and can be divided in three periods: (1) from 0 to 50 h dimer and trimer formation increases while monomer percentage decreases, (2) in the range 50−75 h, the trimer and monomer, respectively, increase and decrease slightly while the dimer remains sensibly constant, and (3) after 75 h, the trimer disappears to the benefit of the monomer and to a lesser extent of the dimer. While period 1 seems to be governed essentially by air oxidation, period 3 can be explained by thiol−disulfide exchange between oxidized trimers and reduced monomers to give oxidized monomers and dimer.

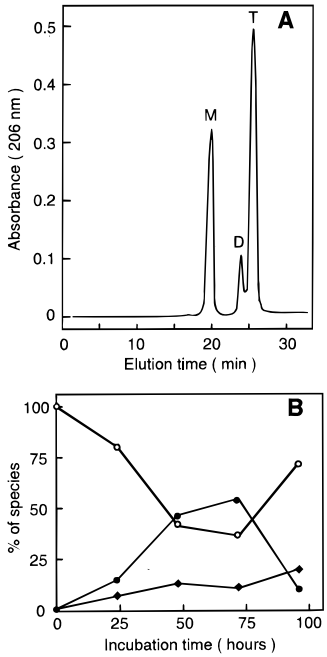


FIGURE 4: Analysis of the association products of the peptide $(GPP^*)_3GYCDPSSCAG$ obtained by air oxidation at −20 °C. (A) The peptide $(GPP^*)_3GYCDPSSCAG$ (1 mg/mL) was incubated in the association buffer, 50 mM Tris-HCl, 100 mM NaCl, pH 8.5, and 30% methanol, for 72 h at −20 °C before analysis by RP-HPLC as described in Figure 1. M, D, and T refer to monomer, dimer, and trimer, respectively, as identified by electrospray mass spectroscopy on a LAPI Sciex spectrometer. (B) Kinetics of disulfide-bonded species formation: open circles, closed squares and closed circles refer to monomer, dimer, and trimer, respectively. At various times, aliquots were removed from the association buffer, acidified, and analyzed by RP-HPLC to quantify the levels of each species.

The oxidized monomer is the most thermodynamically stable molecule among all molecular species present in the medium. Thus, the formation of trimer appears to occur under kinetic control. The relative stability of the dimer during period 2 highlights that this compound is probably an intermediate species in trimer formation. These results show that incubation time is critical to obtain an optimum amount of trimer formation. The highest yield (55%) is obtained after 72 h of incubation where the peptide is preferentially associated into trimer (T in Figure 4A).

The effect of peptide concentration on peptide assembly was also investigated (data not presented). As expected, the rate of trimer formation decreases when the peptide concentration is lower than 1 mg/mL, but it is worth mentioning that even at 0.01 mg/mL peptide self-association into trimer remains efficient. This suggests a high specificity in the trimer association.

As the rate of trimer formation is slow, we investigated the use of metal catalysts known to enhance markedly air oxidation of peptides. Addition of Cu^{2+} increases the rate of trimer formation but does not modify the relative ratio of the various molecular species (not shown). Figure 5 illustrates the effect of various concentrations of Cu^{2+} added in the association buffer on the rate of trimer formation after 15 h of incubation. With a 10 μM concentration of Cu^{2+}, 40% of covalently linked trimers are already formed after only 15 h of incubation while only 10% are obtained in the absence of Cu^{2+}. With 100 μM Cu^{2+}, the rate of oxidation is faster, but in addition to monomers, dimers, and trimers,

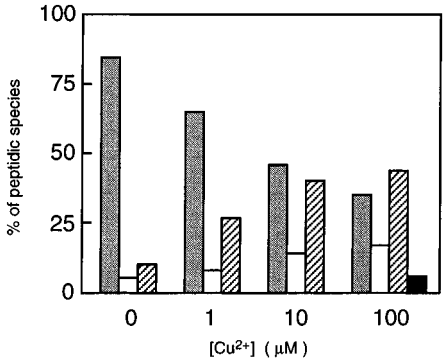
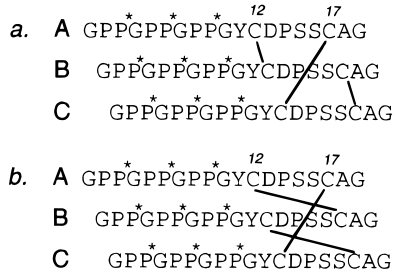


FIGURE 5: Effect of metal catalyst on the association kinetics. The peptide $(GPP^*)_3GYCDPSSCAG$ (1 mg/mL) was incubated in the association buffer, 50 mM Tris-HCl, 100 mM NaCl, pH 8.5, and 30% methanol, for 15 h in the presence of growing concentrations of Cu^{2+} at −20 °C before analysis by RP-HPLC. The percentage of each peptidic species was evaluated from the area of the corresponding peak on the chromatograms. Gray, white, hatched, and black rods refer to monomer, dimer, trimer, and oligomers, respectively.

Scheme 1



other compounds eluting later in RP-HPLC appear. These compounds are higher polymers, indicating that at this concentration of Cu^{2+}, peptide self-assembly becomes less specific.

From the above studies, a standard procedure was established to obtain a high yield of trimer assembly in a relatively short time: the peptide $(GPP^*)_3GYCDPSSCAG$ was incubated at 1 mg/mL in 30% methanol, 10 μM Cu^{2+}, 50 mM Tris-HCl, and 100 mM NaCl, pH 8.5, exposed to air for 15 h at −20 °C before purification. Purified trimer was concentrated by lyophilization, dissolved at about 1 mM in methanol, and stored at −20 °C.

Stability of the Trimer. The purified trimer stored in methanol appeared very stable since almost no monomer and dimer were detected by RP-HPLC, even after several months of storage. Several tests of stability were done in Tris and phosphate aqueous buffer solution at various pHs (4.0−8.0) and temperatures (4−25 °C). After several days of incubation at 1 mg/mL, RP-HPLC analysis of the solutions showed that the trimer did not significantly dissociate. Only the assays performed after 4 days of incubation in the above buffer at 25 °C and pH = 8.0 showed significant decomposition of trimer in monomer (about 5%). These data indicate that the trimer is very stable once purified, i.e., separated from reducing compounds.

Disulfide Bonds. The measurement assays of free thiol groups with DTNB at 412 nm according to Ellman's method (Riddles et al., 1983) did not show any release of thionate ion. This indicates that all six cysteine residues of the trimer are involved in disulfide bonds. Scheme 1 represents the two possible isomers of the $[(GPP^*)_3GYCDPSSCAG]_3$ trimer with three interchain disulfide bonds (solid lines).

In this representation, the three chains are staggered from one amino acid to take into account the formation of a collagenous triple helix at the N-terminal end. The three different positions of each chain are numbered A, B, and C. The first isomer *a* is characterized by one disulfide bond between the C12 of a chain and the C17 of another chain, one disulfide bond between two C12, and one disulfide bond between two C17. Taking into account a collagenous triple helical structure for the GPP* triplets, the three chains are shifted and not spatially equivalent. Consequently, six conformers must be considered with this first arrangement of disulfide bonds (and 36 if the three chains are to be distinguished, for instance, by different primary sequences or different conformations). The second isomer *b* is characterized by three disulfide bonds between the C12 of one chain and the C17 of another chain. Only two conformers with this second arrangement of disulfide bonds must be considered (and 12 if the three chains are to be distinguished).

To determine which type of isomer was formed by peptide self-association, the position of disulfide bonds was investigated by microsequencing. In the 12th cycle of Edman degradation, a release of dehydroalanine was observed reflecting the presence of the first isomer *a*. Indeed, if only isomer *b* of the trimer was present, no release of any PTH amino acid would be observed. It should be mentioned that in the 12th cycle di-PTH-cystine could not be precisely quantified since its retention time is close to that of PTH-tyrosine, released in the preceding cycle. In the 17th cycle of Edman degradation, the amount of dehydroalanine released was about two times that observed in cycle 12, as expected for isomer *a*. However, an accurate quantification was not possible due to (i) the low yield of di-PTH-cystine release, (ii) the partial conversion of di-PTH-cystine into PTH-dehydroalanine which increases with the number of cycles, (iii) traces of PTH-dehydroalanine coming from serine degradation in cycles15 and 16, and (iv) the decrease of yield at each Edman's cycle. Consequently, although the data are consistent with isomer *a* of the trimer, at this stage, the presence of isomer *b* cannot be completely excluded.

Circular Dichroism. The formation of a collagen structure was investigated by circular dichroism. The CD spectrum collected at 1 °C in a phosphate buffer shows a minimum band around 197 nm and a positive band around 228 nm (Figure 6A, bold curve). These are distinguishing features of a coiled-coil triple helix in collagen (Piez & Sherman, 1970). At 40 °C, the magnitude of the minimum at 197 nm decreased and $[\theta]_{228nm}$ comes to zero, indicating the melting of the triple helix (Figure 6A, thin curve). Spectra collected in methanol present similar features (Figure 6B), but the magnitude of the maximum band around 228 nm is much higher than in aqueous solution. The thermal transition curves obtained by monitoring $[\theta]_{228nm}$ as a function of temperature exhibit a well-defined transition at about 30 °C for data collected in methanol (Figure 6C, curve b) and a transition at about 15 °C for data collected in a phosphate buffer (curve a). The higher magnitude of the positive band around 228 nm as well as the higher melting temperature in methanol than in an aqueous buffer indicates a better stability of the collagen-like conformation in methanol. This observation is consistent with previous studies on collagenous peptides, showing that solvents of weaker hydrogen-bonding strength than water provide a more favorable medium to form the interchain hydrogen bonds required for the stabilization

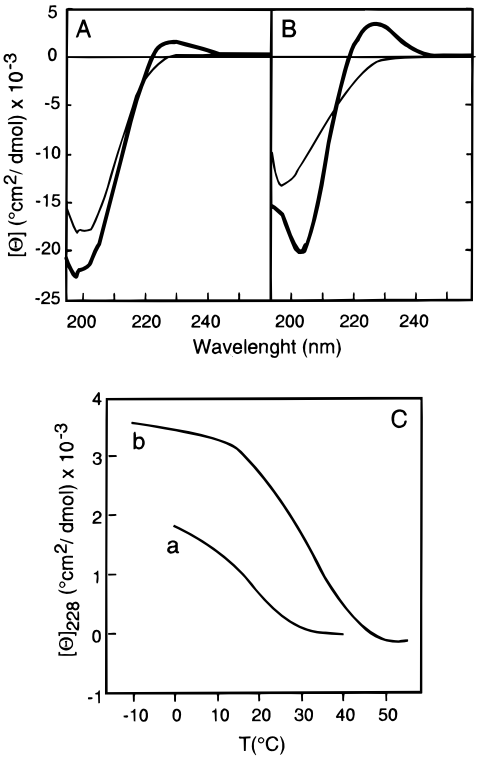


FIGURE 6: Circular dichroism spectra of the trimer $[(GPP^*)_3GYCDPSSCAG]_3$. The spectra were recorded at a concentration of 0.5 mg/mL either in a 50 mM sodium phosphate buffer, pH = 7 (A) at 1 °C (bold curve) and 40 °C (thin curve), or in methanol (B) at −10 °C (bold curve) and 50 °C (thin curve). (C) Thermal transition curves for the trimer in a 50 mM sodium phosphate buffer, pH = 7 (a), and in methanol (b).

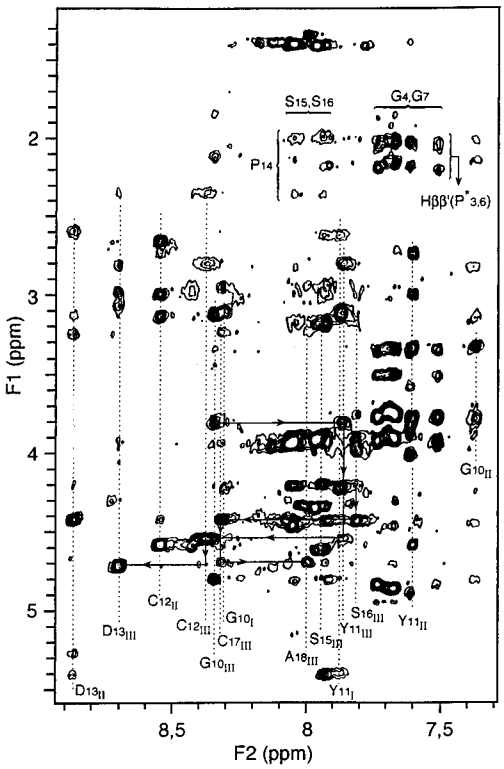


FIGURE 7: NOE correlations of the amide protons with aliphatic protons in the trimer $[(GPP^*)_3GYCDPSSCAG]_3$ (CD_3OH, 10 °C, mixing time 250 ms). Sequential assignment is indicated for G10 to D13 and for S15 to A18 of chain III.

of the triple helical collagen-like conformation (Brown et al., 1972).

NMR Spectroscopy. NMR data obtained on the trimer dissolved either in water, in methanol, or in a mixture of water/methanol confirmed the observations made with CD spectra: at a fixed temperature (typically 10 °C), the signals of 1D and 2D NMR spectra were appreciably broadened when the content of water increased, indicating conformational averaging processes (data not shown). The characterization of the trimeric 3D structure was thus undertaken from spectra in 100% methanol, which presented the best resolution. However, even in this solvent, the spectra showed the presence of broad peaks among thiner ones, as illustrated in the NOESY spectrum of Figure 7. This suggests that conformational motions occur in some parts of the molecule in the intermediate time scale relative to the chemical shifts., i.e., in the millisecond time scale, and that, even in our conditions of solvent and temperature, the structure of the trimer is not fully stabilized. Besides, no exchange cross-peaks were observed in ROESY spectra (200 and 300 ms mixing time); i.e., no slow exchange between conformers in the second time scale was detected.

Examination of NOESY spectra of Figures 7 and 8 shows that the trimer is not a symmetrical molecule. This is particularly well illustrated by the three Tyr of the molecule which give three very distinct sets of aromatic spin systems (Figure 8), of approximately equal intensities as observed on 1D spectrum (not shown). These residues are thus in distinguishable chemical environments, indicating a totally asymmetric assembly of the trimer. Although some minor aromatic spin systems of Tyr can be detected in the background noise of 1D and 2D spectra, the resonances of the three major Tyr rings represent at least 90% of signal intensity. This suggests the presence of one major molecular species for the trimer. Similar observations were made after resonance assignment for other residues of the molecule, which further support these conclusions. NOESY spectra show a large number of cross-peaks, especially in the amide region (figure available as Supporting Information), indicating that this major asymmetric species observed for the trimer is folded.

Assignment Procedure. Difficulties in resonance assignments arose from the presence of three identical chains in the molecule. Moreover, there are repetitive amino acids in each chain (5 Gly, 4 Pro, 3 Hyp, 2 Cys, and 2 Ser) and only six different spin systems among the 57 amino acids of the trimer. Additional difficulties came from the extensive overlaps in some regions of the spectra and from the broadening of some resonances, which led to a severe decrease of the resolution. Finally, a number of less intense cross-peaks likely corresponding to minor isomers were also observed.

Identification of the amino acid spin systems and sequential assignment of $[(GPP^*)_3GYCDPSSCAG]_3$ were achieved using mainly the classical strategy described by Wüthrich (1986). Spin systems were identified at various temperatures (10, 20, and 30 °C) using DQF-COSY and TOCSY spectra and with the help of $^{13}C-^{1}H$ HSQC spectra. Sequential assignments were made on the basis of NOESY spectra collected at 150 and 250 ms mixing times.

At first, hydroxyproline (P*) has the same connectivity pattern as proline (P) but is distinguished from P by the Hγ which is shifted downfield by the OH group. Both types of residues (12 P and 9 P*) gave several sets of overlapping

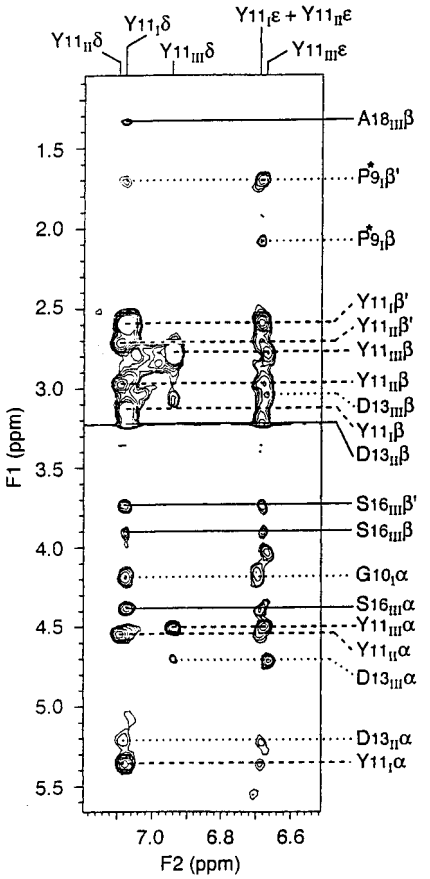


FIGURE 8: NOE correlations of the aromatic protons of tyrosine residues in the trimer $[(GPP^*)_3GYCDPSSCAG]_3$ with aliphatic protons (CD_3OH, 10 °C, mixing time 250 ms). The vertical solid lines at the bottom of the spectrum indicate the aromatic proton resonances of the tyrosine residues of chains I, II, and III. The horizontal lines show interchain (—), sequential (···) and intraresidue (- - -) connectivities.

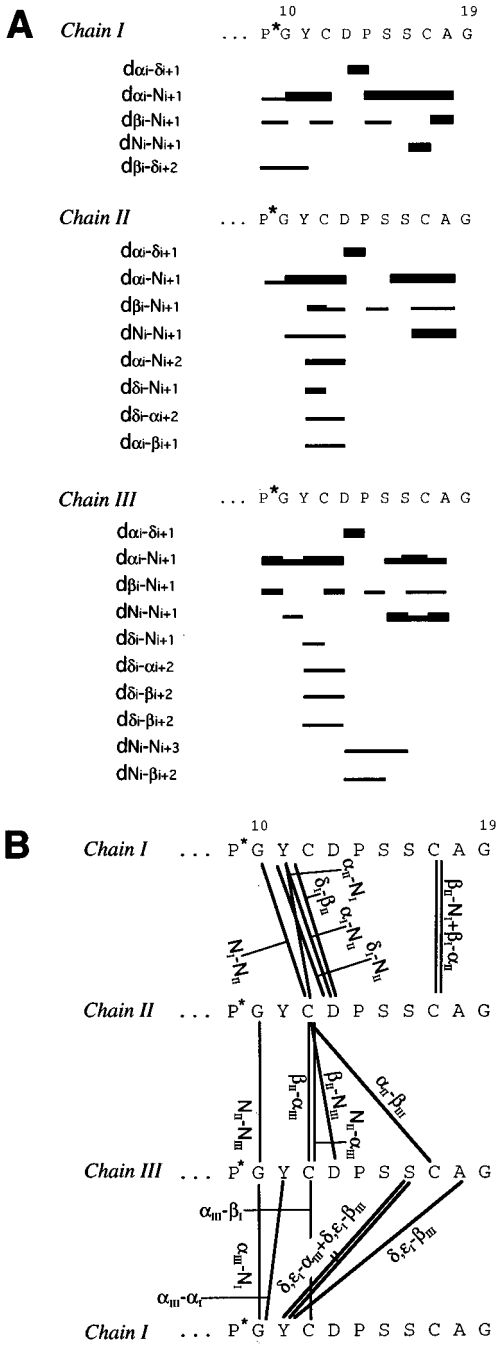


FIGURE 9: Summary of the observed intrachain (A) and interchain (B) NOE connectivities in the trimer $[(GPP^*)_3GYCDPSSCAG]_3$. In panel B, the interchain connectivities between residues are indicated by solid lines. Wide, medium, and narrow bars (panel A) or lines (panel B) indicate respectively strong, medium, and weak NOE intensities.

resonances. However, it is worth mentioning that the H$\delta\delta'$ of the three P14 exhibited three well distinct sets of resonances [as identified by their NOE correlations with their neighboring Hα(D13); see below] at about 0.5 ppm lower field than those of the Pro residues belonging to the triple helix. These well-separated H$\delta\delta'$(P14) allowed the identification of the H$\beta\beta'$(P14) and H$\gamma\gamma'$(P14) on TOCSY spectra. The other spin systems were distinguished from P and P* by the amide proton resonances. Apart from Gly and Ala residues which were respectively characterized by a pair of Hα resonances and by a methyl group, all remaining residues were AMX spin systems. Ser residues differed from Tyr, Asp, and Cys by their Hβ resonances shifted downfield, and Tyr residues were easily identified by strong intraresidue NOE correlations between H$\delta\delta'$ and Hα/H$\beta\beta'$. The identification of the Hβ pair attached to the same carbon was unambiguously achieved with the $^{13}C-^{1}H$ HSQC spectra. These experiments also gave additional evidence for the assignment of the spin systems by the ^{13}C chemical shift (Medvedeva et al., 1993; Wishart et al., 1991, 1995).

For the amino acids located in the noncollagenous part of the trimer (residues 9−18 in each chain), despite the difficulties mentioned above, both the identification of the spin systems and the sequential assignment were successful for the three chains. The three distinct peptidic chains were arbitrarily named I, II, and III (Figure 9). Sequential assignments were made essentially on the basis of the observation of H$\alpha(i)-$NH$(i+1)$ NOE correlations and confirmed, when possible, by H$\beta(i)-$NH$(i+1)$ NOE correlations. For the Asp−Pro bonds, the strong Hα(D13)−H$\delta\delta'$(P14) correlations observed were used for the sequencial assignment (by the way, these correlations indicate a *trans* Asp−Pro peptidic bond for the three chains). For the Pro−Ser bonds, H$\beta\beta'$(P14)−NH(S15) and H$\gamma\gamma'$(P14)−NH(S15) NOE correlations were used because the three Hα(P14) were not unambiguously identified. Attributed DQF-COSY and $^{13}C-^{1}H$ HSQC spectra extracts as well as two tables containing ^{1}H and ^{13}C chemical shifts are available as Supporting Information.

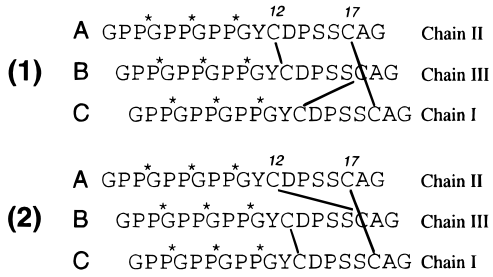
Once the sequential assignment of the three chains from residues P*9 to A18 was complete, NOESY spectra were carefully examined for medium-range ($|i - j| \leq 4$) and long-range ($|i - j| > 4$) intrachain NOEs but also for interchain connectivities. For example, Figure 8 illustrates the different types of NOE correlations (i.e., intraresidue, intrachain, and interchain connectivites) observed between the aromatic protons of Tyr residues and aliphatic protons. It should be mentioned that one *a priori* difficulty of the sequential assignment was the fact that a given $i - i + 1$ NOE cross-peak might result from an interaction between atoms in one chain or between atoms in different chains. In fact, it appeared that the sequential assignment of each chain was not ambiguous since all the interchain NOE correlations but one [Hβ(C12$_{II}$)−NH(D13$_{III}$)] could not be confused with sequential ones and since most of the sequential NOE correlations were intense. Finally, a few remaining minor spin systems were observed in TOCSY and DQF-COSY spectra that could not be sequencially assigned neither between them nor to one of the identified chains of the major trimeric species. Thus, they are likely related to minor species.

For the repetitive sequences GPP*, sequential assignment was not possible due to the extensive overlapping of resonances, especially in the aliphatic region of the spectra. The nine sets of resonances for Pro and Hyp residues could not be individually observed. Although the amide resonances for G4 and G7 could be easily identified by their correlations on NOESY spectra with the side chains of Hyp residues (Figure 7), these residues could not be individually assigned. However, NH−NH cross-peaks were seen on NOESY spectra between these Gly residues. Such NOE correlations may be expected in a triple helical conformation between glycine residues of adjacent chains.

All sequential and medium-range NOEs are summarized in Figure 9A while the observed interchain connectivities are schematically represented in Figure 9B. No intrachain long-range NOEs were observed. Whatever the chain considered, the region around C12 is defined by several interchain NOEs. The C-terminal region around C17 of chain III, which presents a particular conformation as shown below (Figure 11B), is characterized by four "long-range" interchain connectivities involving S16, C17, and A18. On the contrary, the C-terminal ends of chains I and II present only two interchain NOEs between their C17 residues. Interestingly, constraints observed between residues D13$_{III}$ and S15$_{III}$/S16$_{III}$ are close to that obtained for the cyclized monomer (Figure 3). In particular, a NH(i)−NH(i+3) correlation between residues D13 and S16 is observed in both cases. Note that no NOE correlation was observed between the aromatic protons of Tyr residues.

Arrangement of Chains and Disulfide Bonds. The three chains I, II, and III have to be staggered by one residue in their N-terminal end to account for the triple helical structure of the GPP* repeats. According to Scheme 1, the three different positions of the chains were numbered A, B, and C. The relative position of the three chains I, II, and III can be deduced from NOE connectivities. The observation of NOE correlations between G10$_{III}$ and G10$_{I}$ and between G10$_{III}$ and G10$_{II}$ (Figure 9B) suggests that chain III occupies the position B. A NH−NH NOE cross-peak between G10$_{II}$ and a Gly residue of the triple helical region indicates that chain II occupies the position A, as represented in Scheme 2. Moreover, the fact that most of the interchain constraints between chains I and II are of i(I)-i+2(II) type is in agreement with the shift of two amino acids between these chains near the triple helical region (Scheme 2). Although care must be taken due to the proximity of Y11 aromatic rings, the analysis of the chemical shifts for G10 residues supports the above conclusion. Indeed, the amide proton of G10$_{II}$ at 7.38 ppm is close to those of Gly residues located in the triple helical part of the molecule, G4 and G7 (between 7.5 and 7.8 ppm), whereas the amide protons of G10$_{I}$ and G10$_{III}$ are shifted downfield (8.32 and 8.34 ppm, respectively). In the same way, the chemical shifts of G10$_{III}$ α protons are much more similar to those of G4 and G7 than those of G10$_{I}$ and G10$_{III}$. Thus, G10$_{II}$ has a chemical environment close to Gly residues involved in the triple helix structure, in agreement with chain II occupying position A (Scheme 2).

Scheme 2



The observation of two interchain NOEs between C17$_{I}$ and C17$_{II}$ together with the absence of any other interchain NOE in the C-terminal end of chains I and II, highly suggests the presence of a disulfide bond between these two cysteines. This assumption excludes the disulfide bond pattern of isomer *b* in Scheme 1, which is in agreement with the microsequencing results. Spatial proximities between C12$_{I}$, C12$_{II}$, C12$_{III}$, and C17$_{III}$ as shown in Figure 9B do not allow any *a priori* hypothesis about the two other disulfide bonds. The two remaining possible disulfide arrangements (called 1 and 2) are illustrated in Scheme 2. These conclusions were further supported by molecular modeling calculations as reported below.

Molecular Modeling. The self-consistency of the data and the number of observed NOEs suggested that a more quantitative determination of the structure could be made by a simple annealing protocol using the NOE data to provide interproton distance restraints. Although the analysis of the observed NOEs could be made in terms of a single, predominant conformer, caution must be exercised in calculating 3D structures because of the presence of minor species and possible conformational exchange in the intermediate NMR time scale. The distance constraints were thus estimated only as upper limits.

For the triple helical part of the trimer (residues G1 to P*9 in each chain), no NOE constraint could be extracted from NMR spectra. Thus, to model this region of the trimer, we took advantage of the knowledge of the 3D structure of a collagenous triple helix recently solved from X-ray crystal diffraction for the collagen-like peptide (PP*G)$_4$PP*A-(PP*G)$_5$ (Bella et al., 1994). The coordinates of the three N-terminal GPP* triplets of each chain were used to calculate the corresponding distance constraints and dihedral angles in order to generate the triple helical conformation as described in Materials and Methods. Examination of cal-

Table 2: Distribution of Distances (Å) between S Atoms of Cysteine Residues over 100 Structures Obtained after X-PLOR Calculation[a]

	residue					
residue	$C12_I$	$C17_I$	$C12_{II}$	$C17_{II}$	$C12_{III}$	$C17_{III}$
$C12_I$		*12.4 ± 3.1*	*9.1 ± 2.8*	*12.8 ± 3.6*	6.7 ± 1.5	9.5 ± 2.6
$C17_I$	*intrachain*		*13.9 ± 2.7*	**5.2 ± 1.2**	*15.2 ± 2.7*	*12.6 ± 4.2*
$C12_{II}$	*excluded*	*excluded*		*13.2 ± 2.6*	7.2 ± 1.7	7.7 ± 1.9
$C17_{II}$	*excluded*	**sure**	*intrachain*		*15.4 ± 2.8*	*11.9 ± 4.1*
$C12_{III}$	possible	*excluded*	possible	*excluded*		*9.3 ± 2.6*
$C17_{III}$	possible	*excluded*	possible	*excluded*	*intrachain*	

[a] The average values and standard deviations are given above the diagonal. Structure calculations were performed using the X-PLOR program as described in Materials and Methods, without fixing any constraint between S atoms. The bold value is statistically the lowest distance involving $C17_I$ and $C17_{II}$ and thus likely reflects a disulfide bond. Italic values would correspond either to intrachain disulfide bonds or to excluded disulfide arrangements when considering the $C17_I-C17_{II}$ bond. The four remaining values (underlined values) correspond to possible disulfide bonds. These considerations are summarized below the diagonal.

Table 3: Structural Characteristics of the Trimeric Structures Generated by Computer Modeling[a]

	no disulfide[b]	disulfide pattern 1[c]	disulfide pattern 2[d]
NOE violations[e]			
number >0.5 Å	0.2 ± 0.4	0.6 ± 1.0	0.4 ± 0.7
between 0.4 and 0.5 Å	0.5 ± 0.7	0.7 ± 0.9	0.8 ± 0.9
between 0.3 and 0.4 Å	1.7 ± 1.1	1.3± 1.2	1.8 ± 1.1
between 0.2 and 0.3 Å	2.0 ± 1.9	2.2 ± 1.7	2.8 ± 2.3
between 0.1 and 0.2 Å	6.6 ± 3.1	7.0 ± 2.1	7.5 ± 3.1
av max violation (Å)	0.4 ± 0.1	0.5 ± 0.1	0.5 ± 0.1
max consistent violation (Å)	0.9	0.8	0.8
X-PLOR energy (kJ/mol)	−118 ± 45	−74 ± 63	−66 ± 63
covalent geometry			
bond length deviation ($\times 10^{-4}$ Å)	50 ± 5	60 ± 8	60 ± 9
valence angle deviation (deg)	1.5 ± 0.1	1.6 ± 0.2	1.7 ± 0.2
improper angle deviation (deg)	0.3 ± 0.1	0.4 ± 0.1	0.4 ± 0.1
RMSD values (Å)[f]			
on 100 calculated structures			
all three chains	5.64	5.79	5.54
chain I	4.55	4.70	4.44
chain II	3.76	3.56	3.61
chain III	3.69	3.79	3.72
$C12_{II}-C17_{II}$	1.56	1.57	1.66
$C12_{III}-C17_{III}$	1.19	1.21	1.29
on subsets of structures			
$C12_{II}-C17_{II}$	0.93 (45)	0.97 (36)	1.04 (25)
$C12_{III}-C17_{III}$	0.92 (77)	0.71 (65)	0.9 (25)

[a] A total of 100 structures were calculated using the X-PLOR package. [b] No additional distance constraint between S atoms was introduced. [c] The disulfide bonds were fixed according to pattern 1 of Scheme 2, i.e., $C12_I-C17_{III}$, $C17_I-C17_{II}$, and $C12_I-C12_{III}$. [d] The disulfide bonds were fixed according to pattern 2 of Scheme 2, i.e., $C12_I-C17_{III}$, $C17_I-C17_{II}$, and $C12_I-C12_{III}$. [e] The mean number of NOE violations per calculated structure and the corresponding standard deviation are indicated; statistics given over 100 structures. [f] The pairwise RMSDs were calculated either on the whole set of 100 generated structures or on subsets of strucures (threshold of pairwise RMSDs of 1 Å). The number of selected structures is indicated in parentheses.

culated structures showed that the interchain hydrogen bonds between amide protons of glycine residues and carbonyl groups of hydroxyproline residues were established according to the molecular model proposed from X-ray diffraction (Bella et al., 1994).

For the noncollagenous part of the trimer (residues G10 to G19 in each chain), the observed NOEs were converted into 131 distance constraints (37 intraresidue, 61 sequential, 12 medium-range, and 21 interchain). A set of 100 structures was calculated using the X-PLOR package and analyzed in terms of convergence as detailed below. At first, these structures were used to determine the location of the disulfide bonds. In agreement with the above observations, the analysis of the S−S distance distribution (Table 2) yields to the identification of one disulfide bond between $C17_I$ and $C17_{II}$ ($d_{S-S} = 5.2 \pm 1.2$ Å). The existence of this bond excludes all other bonds involving these two cysteines (which in addition exhibit S−S distances >12.6 Å) as well as all bonds which would not be compatible with the presence of three interchain S−S bonds (italic in Table 2). However, two arrangements remain possible for the two other disulfide bonds, i.e., either (1) $C12_{III}-C12_{II}$ and $C17_{III}-C12_I$ or (2) $C12_{III}-C12_I$ and $C17_{III}-C12_{II}$, which correspond to arrangements 1 and 2 of Scheme 2. Indeed, the S−S distances between $C12_I$ and $C12_{III}$ or $C17_{III}$ on the one hand, and between $C12_{II}$ and $C12_{III}$ or $C17_{III}$ on the other hand, are in the same range (underlined values).

As no unambiguous disulfide arrangement could be deduced from the upper analysis, the two possibilities (Scheme 2) were considered in subsequent structure calculations. In both cases, 100 structures were calculated, using additional distance constraints between the pairs of sulfur atoms. The resulting structures as well as those obtained without fixing any disulfide bond did not exhibit significant differences in terms of NOE constraint violations, energy, and RMSD (Table 3). Thus, neither energy or RMSD criteria nor the S−S distance analysis (Table 2) gives evidence to distinguish between the two possible disulfide arrangements of Scheme 2. In other words, no single disulfide pattern accurately represents the experimental NMR

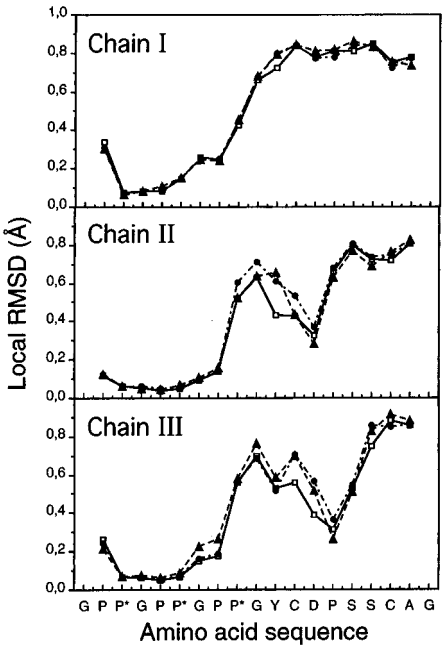


FIGURE 10: Local RMSD plots calculated as described in Materials and Methods on chains I, II, and III. Solid lines and open squares, broken lines and closed circles, and dotted lines and closed triangles refer to calculations performed on 100 structures generated respectively without fixing any disulfide bond, with fixing disulfide pattern 1 of Scheme 2 ($C12_I$−$C17_{III}$, $C17_I$−$C17_{II}$, and $C12_{II}$−$C12_{III}$), and with fixing disulfide pattern 2 of Scheme 2 ($C12_{II}$−$C17_{III}$, $C17_I$−$C17_{II}$, and $C12_I$−$C12_{III}$).

data, and the presence of both disulfide conformers can not be excluded.

Structure comparisons by pairwise RMSD for N, Cα, and C′ backbone atom coordinates showed that, in any case, no satisfying convergence was obtained when considering either the whole trimer or each of the three chains (Table 3). Consequently, it was not possible to get a precise representation of the tridimensional structure of the whole trimer. This low definition is due to the lack of interchain NOE connectivities, mainly in the C-terminal part of chains I and II, while a relatively high number of NOE constraints are available for chain III. Nevertheless, local analogies could be detected when three-residue average RMSDs were calculated along each chain, as illustrated in Figure 10. Besides the collagenous part which is well-defined as expected (local RMSD around 0.1 Å), chain III exhibits low local RMSDs in the region beween C12 and C17, whatever the disulfide pattern considered. This is also observed to a lesser extent for chain II but not for chain I. For the whole set of structures calculated without fixing disulfide bonds, a RMSD of 1.19 Å is obtained for $C12_{III}$ to $C17_{III}$ fragment (Table 3). This indicates that all the structures have a similar folding in this region, which is confirmed by the superimposition of 77 structures selected on the basis of pairwise RMSDs with a threshold of 1 Å (Figure 11B and Table 3). The same analysis performed on chain II also defines the folding of the $C12_{II}$ to $C17_{II}$ fragment, although a larger dispersion of the conformations is observed (Figure 11C, 45 selected structures, Table 3). Similar foldings of these CDPSSC sequences are observed for chains II and III without fixing the disulfide bond arrangement (data not shown).

Further analysis of the CDPSSC fragment of chain III, for which C12 and C17 are disulfide bonded with C12 of chains I and II, showed that its folding is similar to that observed for the cyclized monomer (Figure 11A). In both cases, the proline presents a *trans* conformation, and the distance between Cα(D13) and Cα(S16) is lower than 7 Å, a characteristic distance for a turn as defined by Chou and Fasman (1977). Moreover, the DPSS sequences exhibit hydrogen-bonded turns according to Kabsch−Sander secondary structure rules, as incorporated in the program PROCHEK (Laskowski et al., 1993). Thus, the CDPSSC sequence in chain III adopts a turn-like structure similarly to the cyclized monomer.

In order to summarize the whole structural data obtained in this study, two schematic representations of the FACIT COL1−NC1 junction are illustrated in Figure 11D,E, showing the triple helical conformation of the N-terminal end, the turn-like fold of the chain III DPSS sequence, and the two possible disulfide arrangements (1 and 2 of Scheme 2 in models D and E of Figure 11, respectively). Chains I, II, and III are represented in blue, green, and red, respectively, and the three interchain disulfide bonds appear in black.

DISCUSSION

To further understand the interactions involved in FACIT collagen chain assembly, we analyzed the ability of the peptides GYCDPSSCAG and $(GPP^*)_3$GYCDPSSCAG to self-associate into a three disulfide-bonded trimer and we studied the 3D structure of the $[(GPP^*)_3GYCDPSSCAG]_3$ trimer. This trimer mimics the COL1−NC1 junction of type XIV FACIT collagen.

Assembly of the Trimer. In all experiments performed at room temperature, under conditions of pH that favor the formation of disulfide bonds, cysteine oxidation of both GYCDPSSCAG and $(GPP^*)_3$GYCDPSSCAG peptides mainly leads to the formation of a cyclized monomer via an intramolecular disulfide bond. The structure of the major conformation for the oxidized CDPSSC sequence, obtained from NMR data, is shown in Figure 11A. Residues D, P, and S, present between the two cysteines, belong to the most frequently occurring amino acids in β-turns with residues G and N (Chou & Fasman, 1977). Therefore, it is not surprising that, at room temperature, the major fraction of the peptide tends to fold in such a way that the two cysteine residues are in close contact and readily form an intrachain disulfide bond after oxidation.

The best way to limit monomer cyclization and to obtain a high yield in trimer assembly was to decrease the temperature of incubation down to −20 °C. It is assumed that the decrease of temperature favors the most thermodynamically stable trimer, already present at room tempeature. At this temperature, the stability of the trimeric assembly appears to be sufficiently increased to allow the formation of specific interchain disulfide bonds. However, this is only true with the $(GPP^*)_3$GYCDPSSCAG peptide since only very small amounts of trimer were obtained with the GYCDPSSCAG peptide, even at low incubation temperature. This highlights the fact that the presence of the $(GPP^*)_3$ sequence plays an important role in the trimer formation, as discussed below.

From our experimental data, we conclude that the trimer formation occurs in several steps: (1) interaction of three monomeric peptides stabilized by the formation of a triple helix at low temperature through the $(GPP^*)_3$ sequence, (2)

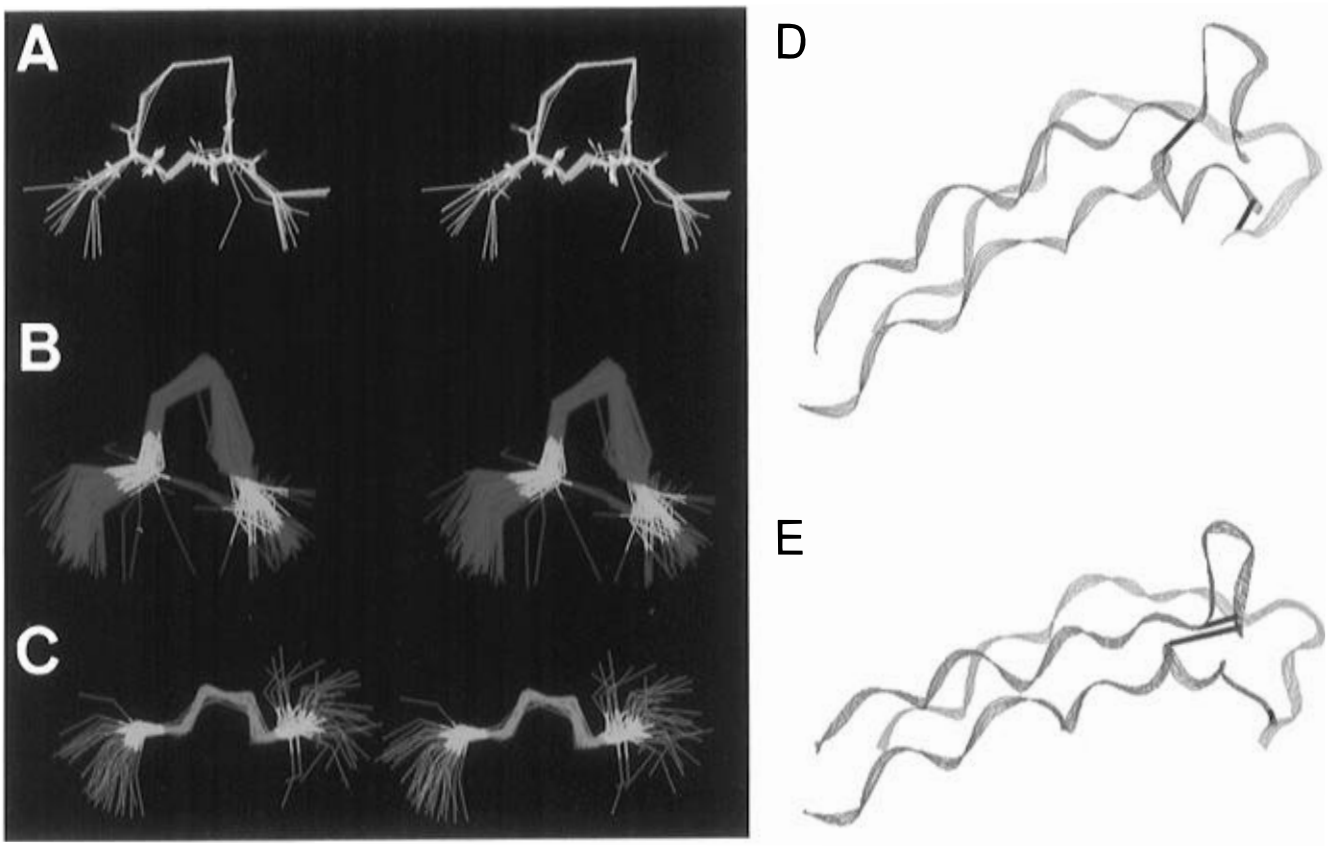


FIGURE 11: Molecular structures for the oxidized peptide GYCDPSSCAG and for the trimer $[(GPP)_3GYCDPSSCAG]_3$, as deduced from NMR data. (A) Stereoview of the backbone superposition of 20 structures for the oxidized peptide GYCDPSSCAG with the *trans* isomery of proline. The disulfide bond appears in yellow. (B and C) Stereoviews of the backbone superposition of 77 (B) and 45 (C) structures of the trimer $[(GPP)_3GYCDPSSCAG]_3$ from residues $C12_{III}$ to $C17_{III}$ (B) and from $C12_{II}$ to $C17_{II}$ (C). (D and E) Structural models of the trimer $[(GPP)_3GYCDPSSCAG]_3$ with disulfide pattern 1 of Scheme 2 (D), i.e., $C12_I-C17_{III}$, $C17_I-C17_{II}$, and $C12_{II}-C12_{III}$, and with disulfide pattern 2 of Scheme 2 (E), i.e., $C12_{II}-C17_{III}$, $C17_I-C17_{II}$, and $C12_I-C12_{III}$. The color code used is the following: blue, green, and red for chains I, II, and III, respectively. The three interchain disulfide bonds are highlighted in black. The N-terminal end of the molecules is in the left of each figure.

formation of the first interchain disulfide bonds by air oxidation, and (3) possible rearrangement of these bonds through thiol and disulfide exchange toward a fully oxidized trimeric conformer. Step 1 is confirmed by the observation that peptides composed of only four GPP* triplets are able to form a stable triple helix at 20 °C (Bruckner et al., 1975). Therefore, it is likely that, at −20 °C, a stable triple helix takes place with only three GPP* triplets. Circular dichroism experiments (Figure 6) clearly show that the purified trimer adopts a triple helical conformation. Moreover, the trimer exhibits much higher melting temperatures in methanol (T_m = 30 °C) and even in water (T_m = 15 °C) when compared to values reported for five GPP* triplets (T_m = 5 °C; Sakaribara et al., 1973). All these observations strongly suggest that both the collagenous part (via a coiled-coil helix formation) and the noncollagenous one (via disulfide bond formation) are involved in the formation and the stabilization of the trimeric molecule. This conclusion is in agreement with the recent results obtained in our group, both *in vitro* with the reassembly of type IX collagen fragments (Labourdette and van der Rest, personal communication) and *in vivo* with the expression of type XII minicollagen (Mazzorana et al., 1995).

When separated from reducing compounds and purified by RP-HPLC, the triple disulfide-bonded trimer is particularly stable, even after long periods of storage. All NMR analyses clearly show that more than 90% of the resonances correspond to one major set of spin systems, *a priori* suggesting that the purified preparation of trimer corresponds to one molecular species. This good structural homogeneity implies that the $(GPP^*)_3GYCDPSSCAG$ peptide contains in itself sufficient information to self-associate stereospecifically into a triple disulfide-bonded trimer.

Structure of the Trimer. The interpretation and resonance assignments of trimer NMR spectra were difficult because (i) the trimer is composed of three identical peptide chains with repeated amino acids in each chain (5 Gly, 4 Pro, 3 Hyp, 2 Cys, and 2 Ser), (ii) the absence of the amide proton for Pro and Hyp, (iii) severe overlaps in some regions of the spectra, and (iv) the broadening of some resonances. Despite these difficulties, the spin systems and NOE correlations were assigned unambiguously for the noncollagenous part of the trimer, i.e., for residues G10 to A18 in each chain. This allowed us to distinguish the three chains and to identify a set of medium-range intrachain and long-range interchain NOEs. The sequential assignment of each chain was not ambiguous since most of the sequential NOE correlations were intense and since all interchain NOE correlations but one could not be confused with sequential ones. A total of 131 distance constraints were actually obtained. As no distance constraint could by extracted for the $(GPP^*)_3$ sequence, a set of 39 interproton distances together with dihedral angles were directly measured from the previously determined 3D collagen triple helix structure (Bella et al., 1994). This set of 170 distance constraints was not sufficient to determine the overall three-dimensional

structure of the trimer. Besides the possible presence of two disulfide conformers (see below), this may be due to slow conformational averaging processes in some parts of the molecule, especially at the C-terminal ends of chains I and II, which result in line broadening and thus in a lack of unambiguously identified NMR constraints. It is very likely that, in the native collagen, the COL1−NC1 junction is stabilized by the presence of the whole NC1 domains.

Although it is not possible to represent the trimer model peptide as a single structure, important structural features of the COL1−NC1 junction have been pointed out in the present NMR study. Concerning the three disulfide bond arrangement, NMR data analysis and molecular modeling clearly show that the cysteine cross-linking is established according to Scheme 2, which is in agreement with the microsequencing results. However, there is no definitive indication in favor of pattern 1 or 2 of Scheme 2, and both conformers may actually be present in our preparation of the synthetic model trimer. Nevertheless, six among the eight possible disulfide isomers could be excluded. In particular, the symetrical disulfide pattern (isomer *b* of Scheme 1) has been unequivocally eliminated. Interestingly, the disulfide bond arrangement previously proposed by Bruckner et al. (1979) for two adjacent cysteine residues located at the junction of the long triple helix and the C-terminal noncollagenous domain of type III collagen also corresponds to isomer *a* of Scheme 1. Bruckner's model differs, however, from ours by the relative position of the three chains. Conversely, the arrangement reported in the C-propeptide of type I procollagen where the two cysteines are separated in each chain by eight amino acids corresponds to isomer *b* of Scheme 1 (Koivu, 1987). These divergences suggest that the arrangement of three disulfide bonds linking three collagenous chains is correlated to the number, and probably the nature, of amino acids located between the two cysteines.

The second main feature of the trimer structure is the particular folding of chain III (in red on Figure 11) into a turn-like conformation, whatever the disulfide pattern considered. The C12 and C17 of chain III form two disulfide bonds with the C12 of the two other chains, conferring a very asymmetric structure to the molecule. Chain III seems to lock the structure by the disulfide bonds between its C17 and the C12 of chain I or II (Figure 11D,E). It is worth mentioning that the NOEs observed for this "locking chain" between residues D13 and S15/S16 are similar to that observed for the cyclic monomer (Figure 3). Indeed, examination of the 3D structures shows that the DPSS sequence has a similar conformation in the locking chain III with interchain disulfide bonds and in the cyclized monomer with an intrachain disulfide bond. Thus, the DPSS sequence presents a high propensity to fold into a turn-like conformation which is very favorable to the formation of interchain disulfide bonds in trimeric assembly. As the COL1−NC1 junction exhibits a strong primary structure homology in all FACITs, the sequences of the four amino acids located between the two cysteine residues should present the same folding properties.

There is no definitive evidence that the structural features reported here for the FACIT COL1−NC1 junction are present in the native molecule. The *in vitro* conditions for peptide self-association are far from cellular ones, where the assembly occurs with the whole collagen chain and where other proteins are likely involved in the trimeric assembly (Nagata, 1996). Nevertheless, the validity of our findings is strongly supported by the specificity of $(GPP^*)_3$GYCDPSSCAG peptide self-association, the high yield, and the stability of the trimer formed. Moreover, the structure of the CDPSSC sequence for the locking chain in the trimeric model is close to the thermodynamically favored cyclized one in the oxidized monomer, despite the fact that the two cysteines are involved in interchain disulfide bonds. In conclusion, the models proposed here (Figure 11D,E) are good representations of the FACIT COL1−NC1 junction. Further works are needed to achieve the three-dimensionnal structure of the COL1−NC1 junction by studying the assembly of peptides containing the whole NC1 domain, either with homotrimeric or heterotrimeric molecules.

ACKNOWLEDGMENT

We thank Dr. L. Labourdette for his advice in biochemical experiments as well as for helpful discussions, J. Bernillon and Dr. J. Wallach for peptide synthesis, Dr. M. M. Boutillon for the microsequencing experiments, Dr. Eric Forest for mass spectroscopy analysis, Dr. C. Ebel for advice in CD techniques, Dr. Y.-S. Yang for his helpful criticisms for NMR experiments, and C. C. Van Herrewege for the artwork.

SUPPORTING INFORMATION AVAILABLE

One table containing the proton chemical shifts of the assigned residues for the various conformers of the peptide GYCDPSSCAG (linear and cyclic, with *cis* or *trans* proline), two tables containing the proton and carbon chemical shifts of the assigned residues for chains I, II, and III of the GPP*GPP*GPP*GYCDPSSCAG trimer, one figure showing the NH−Hα region of a DQF-COSY spectrum, one figure showing a $^1H-^{13}C$ HSQC spectrum, and one figure showing the amide region of a NOESY spectrum for the GPP*GPP*GPP*GYCDPSSCAG trimer (6 pages). Ordering information is given on any current masthead page.

REFERENCES

Abe, N., Hidekatsu, H., Inoue, H., & Ninomiya, Y. (1994) *Biochim. Biophys. Acta 1024*, 61−67.
Bax, A., & Davis, D. G. (1985a) *J. Magn. Reson. 63*, 207−213.
Bax, A., & Davis, D. G. (1985b) *J. Magn. Reson. 65*, 355−360.
Beaven, G. H., & Holidays, E. R. (1952) *Adv. Protein Chem. 7*, 319−386.
Bella, J., Eaton, M., Brodsky, B., & Berman, H. M. (1994) *Science 266*, 75−81.
Berndt, K. D., Güntert, P., & Wüthrich, K. (1993) *J. Mol. Biol. 234*, 735−750.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.
Brewton, R. G., Ouspenskaia, M. V., van der Rest, M., & Mayne, R. (1992) *Eur. J. Biochem. 205*, 443−449.
Brown, F. R., Di Corato, A., Lorenzi, G. P., & Blout, E. R. (1972) *J. Mol. Biol. 63*, 85−99.
Bruckner, P., Rutschman, B., & Engel, J. (1975) *Helv. Chim. Acta 58*, 1276−1287.
Bruckner, P., Bächinger, H. P., Timpl, R., & Engel, J. (1978) *Eur. J. Biochem. 90*, 595−603.
Brünger, A. T. (1992) *X-PLOR, a system for crystallography and NMR*, Yale University Press, New Haven, CT.
Chou, P. Y., & Fasman, G. D. (1977) *J. Mol. Biol. 115*, 135−175.
Creighton, T. E. (1984) *Methods Enzymol. 107*, 305−329.
Deber, C. M., Bovey, F. A., Carver, J. P., & Blout, E. R. (1970) *J. Am. Chem. Soc. 21*, 6191−6198.

Dublet, B., & van der Rest, M. (1987) *J. Biol. Chem. 262*, 17724−17727.
Dublet, B., & van der Rest, M. (1991) *J. Biol. Chem. 266*, 6853−6858.
Geourjon, C., & Deléage, G. (1995) *J. Mol. Graphics 13*, 209−212.
Gordon, M. K., Gerecke, D. R., Dublet, B., van der Rest, M., & Olsen, B. R. (1989) *J. Biol. Chem. 264*, 19772−19778.
Gordon, M. K., Castagnola, P., Dublet, B., Lisenmayer, T. F., van der Rest, M., Mayne, R., & Olsen, B. R. (1991) *Eur. J. Biochem. 201*, 333−338.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 7*, 4546−4553.
Kimura, T., Mattei, M.-G., Stevens, J., Goldring, M., Ninomiya, Y., & Olsen, B. R. (1989) *Eur. J. Biochem. 179*, 71−78.
Koivu, J., & Myllylä, R. (1987) *J. Biol. Chem. 262*, 6159−6164.
Kumar, A., Ernst, R. R., & Wüthrich, K (1980) *Biochem. Biophys. Res. Commun. 95*, 1−6.
Labourdette, L., & van der Rest, M. (1993) *FEBS Lett. 320*, 211−214.
Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
Lee, K. K., Black, J. A., & Hodges, R. S. (1991) in *High-Performance Liquid Chromatography of Peptides and Proteins: separation, analysis and conformation* (Mant, C. T., & Hodges, R. S., Eds.) pp 389−398, CRC Press, Boca Raton, FL.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989) *J. Magn. Reson. 85*, 393−399.
Marti, T., Rösseler, S. J., Titani, K., & Walsh, K. A. (1987) *Biochemistry 26*, 8099−8109.
Mayne, R., & Brewton, R. G. (1993) *Curr. Opin. Cell Biol. 5*, 883−890.
Mazzorana, M., Gruffat, H., Sergeant, A., & van der Rest, M. (1993) *J. Biol. Chem. 268*, 3029−3032.
Mazzorana, M., Giry-Lozinguez, C., & van der Rest, M. (1995) *Matrix Biol. 14*, 583−588.
Medvedeva, S., Simorre, J.-P., Brutscher, B., Guerlesquin, F., & Marion, D. (1993) *FEBS Lett. 333*, 251−256.
Muragaki, Y., Kimura, T., Ninomiya, Y., & Olsen, B. O. (1990) *Eur. J. Biochem. 192*, 703−708.
Myers, J. C., Yang, H., D'Ippolito, J. A., Presente, A., Miller, M. K., & Dio, A. S. (1994) *J. Biol. Chem. 269*, 18549−18557.
Nagata, K. (1996) *Trends Biochem. Sci. 21*, 23−26.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988a) *FEBS Lett. 229*, 317−324.
Nilges, M., Gronenborn, A. M., Brunger, A. T., & Clore, G. M. (1988b) *Protein Eng. 2*, 27−38.
Ninomiya, Y., & Olsen, B. R. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3014−3018.
Ninomiya, Y., van der Rest, M., Mayne, R., Lozano, G., & Olsen, B. R. (1985) *Biochemistry 24*, 4223−4229.
Pan, T., Zhang, R., Mattei, M.-G., Timpl, R., & Chu, M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6565−6569.
Piez, K. A., & Sherman, M. R. (1970) *Biochemistry 9*, 4129−4133.
Rance, M., Sorensen, O. W., Bodenhausen, G.,Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.
Riddles, P. W., Blakeley, R. L., & Zerner, B. (1983) *Methods Enzymol. 91*, 49−61.
Ruiz-Cabello, J., Vuister, G. W., Moonen, C. T., van Gelderen, P., Cohen, J. S., & van Zilj, P. C. (1992) *J. Magn. Reson. 100*, 282−302.
Sakaribara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y., & Prockop, D. J. (1973) *Biochim. Biophys. Acta 303*, 198−202.
Thornton, J. M. (1981) *J. Mol. Biol. 151*, 261−287.
van der Rest, M., & Mayne, R. (1987) in *Type IX collagen. Structure and function of the collagen types* (Mayne, R., Burgeson, R. E., Eds.) pp 195−221, Academic Press, Orlando, FL.
van der Rest, M., & Garrone, R. (1991) *FASEB J. 5*, 2814−2823.
Wetlaufer, D. B (1984) *Methods Enzymol. 107*, 301−304.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311−333.
Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., & Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
Wüthrich, K. (1986) in *NMR of proteins and nucleic acids*, John Wiley & sons, New York.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949−961.

BI952666P