The Self-assembly of a Mini-fibril with Axial Periodicity from a Designed Collagen-mimetic Triple Helix*

Received for publication, May 19, 2014, and in revised form, February 9, 2015 Published, JBC Papers in Press, February 11, 2015, DOI 10.1074/jbc.M113.542241

Parminder Jeet Kaur1, Rebecca Strawn1, Hanying Bai, Ke Xu2, Gabriel Ordas3, Hiroshi Matsui, and Yujia Xu4

From the Department of Chemistry, Hunter College of City University of New York, New York, New York 10065

Background: Functional collagen fibrils are formed through the self-assembly of collagen triple helices. Results: A designed triple helical peptide self-assembles into collagen-like mini-fibrils. Conclusion: The sequence of the triple helix alone is sufficient to “code” for the axial, periodic structure of the mini-fibrils. Significance: The work demonstrates an approach to achieve collagen-like fibrils through the self-assembly of designed triple helices.

In this work we describe the self-assembly of a collagen-like periodic mini-fibril from a recombinant triple helix. The triple helix, designated Col108, is expressed in Escherichia coli using an artificial gene and consists of a 378-residue triple helix domain organized into three pseudo-repeating sequence units. The peptide forms a stable triple helix with a melting temperature of 41 °C. Upon increases of pH and temperature, Col108 self-assembles in solution into smooth mini-fibrils with the cross-striated banding pattern typical of fibrillar collagens. The banding pattern is characterized by an axially repeating feature of ~35 nm as observed by transmission electron microscopy and atomic force microscopy. Both the negatively stained and the positively stained transmission electron microscopy patterns of the Col108 mini-fibrils are consistent with a staggered arrangement of triple helices having a staggering value of 123 residues, a value closely connected to the size of one repeat sequence unit. A mechanism is proposed for the mini-fibril formation of Col108 in which the axial periodicity is instigated by the built-in sequence periodicity and stabilized by the optimized interactions between the triple helices in a 1-unit staggered arrangement. Lacking hydroxyproline residues and telopeptides, two factors implicated in the fibrillogenesis of native collagen, the Col108 mini-fibrils demonstrate that sequence features of the triple helical domain alone are sufficient to “code” for axially repeating periodicity of fibrils. To our knowledge, Col108 is the first designed triple helix to self-assemble into periodic fibrils and offers a unique opportunity to unravel the specific molecular interactions of collagen fibrillogenesis.

Collagen is the major component of the extracellular matrix and is involved in a wide range of cellular functions during tissue development and action. At the core of the diverse functions of this versatile protein is its remarkable ability to form supramolecular structures through self-assembly and/or through interactions with other biomolecules. One typical example is the fibrillogenesis of fibrillar collagens including collagen types I, II, and III of the connective tissues. The fibrillogenesis involves the lateral association of collagen triple helices, the structural unit of collagen, to form long, smooth fibrils with a characteristic 67-nm axial structural feature known as the D-periodicity. Fibrillogenesis of collagen in tissues is a complex process involving other macromolecules; the fibril formation itself, however, is a self-assembly process proceeding from the self-association of the triple helix (1–3). Although many of the structural details of both the triple helix and the collagen fibrils have been elucidated, the molecular recognition mechanisms of the self-assembly process remain poorly understood. Efforts at producing collagen-like, self-assembled fibrils through protein design have not been successful.

The molecular recognition process of the self-assembly of collagen triple helix entails a specific 67 nm staggering at the ends of the associating triple helices (4–6). A triple helix consists of three polypeptide chains having a characteristic Gly-X-Y repeating sequence wrapped around a common axis. The Gly residues at every third position are tightly packed at the center of the helix, whereas the side chains of the X and Y residues are largely exposed to solvent and are directly involved in molecular interactions during the fibril formation (7). The linear, rope-like triple helix has a uniform backbone conformation with an axial rise of ~0.8–0.9 nm for each Gly-X-Y repeat. Fibrillar collagens often have more than 1000 amino acid residues (per polypeptide chain) in a noninterrupted Gly-X-Y repeating arrangement forming a triple helix ~300 nm in length. As the result, each triple helix encompasses ~4.4 67-nm D-periods. The fractional D-period leads to the formation of a 0.4 D overlap zone and a 0.6 D gap zone in the laterally packed fibrils. This alternation of overlap and gap zones on the surface of the fibrils manifests itself as the well known, alternating dark-light striat patterns of collagen fibrils as observed by electron microscope.

Structural studies combining electron microscopy and x-ray fibril diffraction have linked the 67 nm D-periodicity of collagen fibrils to a specific 234-residue staggering of the triple helices during self-association (6, 8, 9). Hulmes et al. (10) were

* This work was supported by National Science Foundation Grant CHE 1022120 (to Y. X.). This work was also supported in part by National Institutes of Health Grant RR003037.
1 Both authors contributed equally to this work.
2 Present address: Cardiovascular Research Foundation, 111 East 59 St., New York, NY 10022.
3 Present address: St. George University, Grenada, West Indies.
4 To whom correspondence should be addressed. Dept. of Chemistry, Hunter College of CUNY, 695 Park Ave., New York, NY 10065. Tel.: 212-772-4310; E-mail: Yujia.xu@hunter.cuny.edu.
Mini-fibril with Axial Periodicity from a Designed Peptide

among the first to connect the 234-residue staggering of the triple helices to the sequence periodicity for type I collagen. Using a computer-aided sequence alignment approach, they demonstrated that by a mutual staggering of multiples of 234 residues both the hydrophobic and electrostatic interactions between associating triple helices will be maximized. However, the identification of this 234-residue repeating unit lacked much of the structurally specific information for collagen that the heptad repeat later revealed for the α-helix coiled-coil. The repeating pattern of the residues among the 4.4 D periods is not evident; no specific periodic clusters of charged or hydrophobic residues have been linked to this periodicity. Subsequent fibrillogenesis studies of collagen under in vitro and/or in vivo conditions have implicated the possible involvement of other factors during the assembly and growth of the D-periodic fibrils. Some studies considered the two short stretches of peptide at the N and C terminus of the triple helix domain of collagen, known, respectively, as the N- and C-telopeptide, to be essential for anchoring the ends of the neighboring triple helices during fibrillogenesis (11, 12); others focused on the critical role of the sequences of natural fibrillar collagens, the critical roles of each of the effectors are difficult to delineate, and their specific effects in fibrillogenesis are difficult to evaluate.

Collagen is also a highly desirable biomaterial because of its remarkable ability to anchor macromolecules and to support cell adhesion and differentiation. To date, self-assembled structures have often relied on the introduction of chemical moieties at the ends of the triple helices, and the assemblies have lacked specific structural features (13, 15–19). The one exception is the D-periodic microfibers formed by the self-assembly of a 36-residue synthetic collagen-mimetic peptide reported by Rele et al. (20). Although the microfiber revealed a regular banding pattern under electron microscope, the 18-nm D-period of the microfibers is larger than the maximum possible end to end distance of 10–12 nm predicted for a 36-residue (per single peptide chain) triple helix. Furthermore, the data indicated that the microfibers were formed only after the triple helix had been thermally unfolded. Thus, the D-periodicity of the microfibers cannot be due to the staggered lateral packing of the triple helix; it most likely reflects a self-assembly process different from that of collagen fibrils.

Here, we report our study on the self-assembly of a recombinant collagen triple helix, designated Col108. The recombinant Col108 is expressed in Escherichia coli and consists of a stretch of 378 residues with a noninterrupted Gly–X–Y repeating sequence. The self-assembled mini-fibrils have a clearly defined 35-nm axial periodicity, and each fully folded Col108 comprises 3.3 such periods. A model is proposed that predicts that the underlying mechanism of the self-assembly of Col108 is driven by the maximization of the triple interhelical interactions modulated by the tandem arrangement of the sequence units. This work accentuates the notion that long range periodicity in a sequence is sufficient to promote periodic self-assembly of a triple helix.

**EXPERIMENTAL PROCEDURES**

*The Expression and the Purification of Col108—*The expression plasmid of Col108 was built on the F877 plasmid used for a previous study (21). The recombinant gene of Col108 was constructed by ligating three synthesized genes each bracketed by restriction enzyme sites (see Fig. 1). All genes were acquired from GenScript Corp and the codons optimized for bacterial expression. Additional bases were inserted in conjunction with the restriction enzyme sites to maintain the Gly–X–Y repeating sequences of the gene product: codon sequence GCC↓TCTAGA for the XbaI restriction site and GGTAC↓CCCG for the KpnI site. The three genes were restricted and ligated to form the final Col108 plasmid insert having BamHI and EcoRI restriction sites at the 5′ and 3′ ends, respectively. The insert was introduced into a modified pET32a(+) plasmid (21, 22).

The resulting gene product is a 52-kDa fusion protein with a His-tagged thioredoxin attached to the N terminus of Col108, separated by a thrombin cleavage site.

*E. coli* strain JM109(DE3) (Promega) was used as the host cell for expression. The transformed cells were cultured in LB/Amp medium at 37 °C, induced with 0.1 mM isopropyl-β-D-thiogalactoside at an optical density of 0.4, and allowed to grow overnight at 25 °C. The cells were lysed by sonication (Vibra cell): three 60-s pulses; duty cycle, 30%; and microtip limit, 4. The His-tagged fusion protein was purified using nickel-nitritol tric-acetic acid metal affinity resin (Qiagen) following the standard procedures and buffer recipes from Talon® metal affinity resin user’s manual (Clontech), except 10 mM imidazole was added to the wash buffer to reduce nonspecific binding, and 300 mM imidazole was used (in the wash buffer) for elution. After elution, the protein was dialyzed against wash buffer (without imidazole) overnight at 4 °C to remove imidazole. Enzymatic cleavage of the fusion protein was achieved by adding thrombin from human plasma (Sigma) and incubating overnight at 4 °C. The digest was further purified by HPLC using a gradient of acetonitrile and 0.1% aqueous TFA; the Col108 peptide was eluted at 50% acetonitrile. The final Col108 peptide was stored as lyophilized powder until use. SDS-PAGE (10% gel, Coomasie Blue staining) was used to check the purity of the sample. A combination of 2.9 mM β-mercaptoethanol and 4 mM DTT was used for the reducing condition. Samples were sent to the Proteomics Resource Center of Rockefeller University for in-gel digestion followed by MSMS sequencing.

For stock solutions, Col108 powder was dissolved in 5 mM acetic acid (HAc), pH 4.0, at a concentration of ~4 mg/ml or in TES buffer (30 mM TES, 30 mM NaHPO₄, and 67.5 mM NaCl, pH 7.4) at a concentration of ~1 mg/ml. The concentration was estimated using an absorbance of 0.23 for a 1 mg/ml solution (molar extinction coefficient of 9.1 × 10³, 1-cm optical path) at 280 nm calculated via Protparam. The CD (Aviv Biomedical Spectrometer model 202-01) wavelength scans were conducted at 4 °C between 190 and 300 nm on 0.2 mg/ml Col108 samples in the corresponding buffers. The temperature melt experiments were conducted on 1 mg/ml samples of

---

5 The abbreviations used are: Hyp or O, 4-R-hydroxyproline; HAc, acetic acid; TES, Na-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid; TEM, transmission electron microscopy; AFM, atomic force microscopy.
Col108, monitored at a wavelength of 225 nm from 4 °C to 65 °C with an equilibration time of 2 min at each temperature (equivalent heating rate: 0.3 °C/min). All samples had been equilibrated at 4 °C for up to 10 days to ensure triple helix formation and were further equilibrated for ~2 days after each dilution before any CD data were taken.

**Fibrillogenesis**—For fibrillogenesis, samples at 1 mg/ml in HAc buffer at 4 °C were mixed with an equal volume of precooled, double strength neutralization buffer (60 mM TES, 60 mM Na2HPO4, and 135 mM NaCl, pH 7.4) to raise the pH. The samples were then transferred to a water bath set at the desired temperature (37 or 26 °C). The final concentration of Col108 was 0.5 mg/ml; the final composition of the TES buffer was 2.5 mM acetic acid, 30 mM TES, 30 mM Na2HPO4, and 67.5 mM NaCl, pH 7.4 (ionic strength I = 0.09).

**Electron Microscope Sample Preparation**—Col108 samples were prepared on 400-mesh Formvar carbon-coated copper grids at 6 and 24 h after fibrillogenesis initiation. 3 μl of incubated sample were allowed to sit on the grid for 100 s; the grids were then rinsed with deionized water for 5 s. The grids were stained with 3 μl of 1% sodium phosphotungstate for 100 s and then rinsed again with deionized water for 5 s. The grids were air-dried before being examined using a Jeol 2100 electron microscope (Jeol Inc.) with the exception of the images in Fig. 3 (B and C), which were taken at the Electron Microscopy and Histology Core Facility of Weill Cornell Medical College. The staining process is difficult to control. Approximately one-third of the time it resulted in the negatively stained mini-fibrils, and the other times the mini-fibrils were positively stained. Occasionally both negatively stained and positively stained mini-fibrils were observed on the same grid.

The axial periodicity was determined by manually measuring the length of a pair of adjacent dark and light bands of negatively stained mini-fibrils, and the other times the mini-fibrils were positively stained. Occasionally both negatively stained and positively stained mini-fibrils were observed on the same grid.

The axial periodicity was determined by manually measuring the distance between the peaks of the phase retrace plots (see Fig. 4), and the histogram of the periodicity was constructed using the FREQUENCY tool of Microsoft Excel based on 40 measurements.

**Analytical Ultracentrifugation**—Analytical ultracentrifuge (Beckman Coulter ProteomeLab XL-I) sedimentation equilibrium trials were run with short columns (3 mm) at speeds of 12,000, 16,000, 20,000, 24,000, and 30,000 rpm at 37 °C. The loading concentrations were 0.05, 0.1, 0.2, and 0.5 mg/ml in 5 mM acetic acid or in TES buffer after 6 h of incubation at 37 °C. The interference data were edited and fit using WinReed and WinNonlinV1.060, respectively. The data were fit using a single-species model as a report on the average apparent molecular mass. The specific volume (V-bar) for Col108 was determined to be 0.7212 using the program Sednterp. The densities of the HAc buffer and the TES buffer were calculated to be 0.99827 and 1.00615, respectively, using the same program.

**Interaction Calculations Examining Periodicity**—The interaction curve of Col108 was calculated according to Hulmes et al. (10). An in-house Perl script was coded that compared two identical amino acid sequences as one sequence was held static and the other was moved laterally by an amino acid shift value of one for each calculation cycle. At each shift value, directly opposed amino acids and the immediately flanking neighbors in the two chains were compared. For each hydrophobic coupling and each electrostatic interaction, a value of one was assigned, and the total interaction value was reported. Interaction values were normalized by the total number of possible interactions at each calculation cycle. Hydrophobic amino acids considered were Val, Met, Ile, Leu, Phe, and Pro; positive amino acids were Lys and Arg; and negative amino acids were Asp and Glu. Calculations were run with both parallel and antiparallel chain alignment. The residue of Pro is usually considered to have a weak hydrophobic interaction; it was included in the calculation because Pro accounts for more than 20% of the X and Y residues in Col108. When not considering Pro to be hydrophobic, the absolute intensity of the peaks was reduced, but the overall periodic features of the curves were not affected.

The foldon adopts a tightly folded β-barrel conformation according to the crystal structure of foldon-(Gly-Pro-Pro)10 (24). Residues at the core were considered sequestered from solvent exposure and unlikely to be involved in molecular interactions during the self-assembly. Thus, only the surface residues Gly-Tyr-Ile-Pro-Glu-Ala-Pro-Arg-Asp-Asp-Gly-Glu-Trp lying along a linear trajectory extending from the triple helix, as determined from visual examination of a crystal structure of the T4 foldon domain (Protein Data Bank code 1RFO), were included for consideration.

**RESULTS**

**The Triple Helical Peptide Col108**—The sequence architecture of the Col108 molecule is shown in Fig. 1. A single Col108 chain has a total of 417 amino acid residues comprised of a 378-residue triple helical domain with a noninterrupted Gly-X-Y repeating sequence and a C-terminal foldon domain included to serve as the nucleation site for the folding of the triple helix (24, 25). The triple helical domain is designed to have three pseudo-repeating sequence units designated U1, U2, and U3, respectively, and a C-terminal (Gly-Pro-Pro)₄ sequence included to mimic the imino acid-rich (Gly-Pro-
Mini-fibril with Axial Periodicity from a Designed Peptide

The gene construct and the sequence architecture of Col108. The restriction enzyme sites used to build the gene are shown in gray boxes. The amino acid sequence of the Col domain is shown in detail, where the numbers indicate the residue positions of the four pieces used from human α1 chain of type I collagen. The residues introduced by restriction enzyme sites are in italic and are underlined. The three repeating sequence units of Col108 are marked by double-headed arrows. The sequence of the foldon domain is not shown because of space limitations.

Hyp$_Y$ nucleation sequence of fibrillar collagens to ensure triple helix formation with correct chain register (26). Each of the three pseudo-identical sequence units of Col108 contains a Col domain: a 108-residue segment (hence the name Col108) consisting of residues 242–256, 296–322, 434–478, and 515–535 of the α1 chain of human type I collagen, and an N-terminal (Gly-Pro-Pro)$_4$ repeating sequence for added conformational rigidity; the U2 unit has two additional Gly-$X$-$Y$ triplets flanking the Col domain introduced by the restriction enzyme sites used for the synthesis of the Col108 gene. The 36 Gly-$X$-$Y$ triplets of the Col domain were chosen for their relatively high helix propensity (27). Additionally, Gly-Pro-Cys-Cys sequences were inserted at the N and C termini of the 378-residue triple helical domain to covalently link three constituent chains through terminal Cys knots (28). In summary, the Col108 peptide is comprised of three pseudo-identical sequence units, each having 120 residues (126 residues for U2), consisting of one (Gly-Pro-Pro)$_4$ sequence and one Col domain. The repeating sequence units confer sequence periodicity to the triple helix domain of Col108: the sequence unit is repeated in tandem three times.

As a product of bacterial expression, Col108 has no hydroxyproline, yet it forms stable triple helices upon incubation (~2 days) in 5 mM HAc (pH 4) at 4 °C. The CD spectrum of Col108 is characteristic of a collagen triple helix having a positive peak at 225 nm and a deep, negative peak at 197 nm (Fig. 2A). The ratio of the positive intensity over the negative intensity, also known as the Rpn, of Col108 is ~0.067. This value is lower than those observed for (Pro-Hyp-Gly)$_n$-based peptides having a high content of Hyp (29, 30) but is in good agreement with that of peptide (Gly-Pro-Pro)$_{10}$ (Gly-Pro-Pro)$_{10}$ foldon (22, 26), and other triple helical peptides with high (Gly-Pro-Pro) content (31). The Col108 triple helix has an apparent melting temperature of ~41 °C in HAc buffer (Fig. 2B), despite lacking Hyp residues in the $Y$ positions. The foldon domain, the higher than usual content of charged residues in the Col domain, the regular (Gly-Pro-Pro)$_4$ inserts, and the Cys knots at both the N and C terminus may all contribute to the thermal stability.

The Self-association of Col108—The Col108 triple helices self-associate into large aggregates upon mixing with an equal volume of cold, double-strength neutralization buffer (pH 7.4) and incubating at 26 or 37 °C. The self-assembly in solution can be followed using analytical ultracentrifugation and is demonstrated by the drastically increasing apparent molecular mass average of the samples. The molecular mass average of Col108 in HAc is 114 (~10.9) kDa, a value in good agreement with the expected molecular mass of 117 kDa calculated for a single triple helix of Col108 from the amino acid sequence. In comparison, the molecular mass average of Col108 in TES buffer is 462 kDa after 6 h of incubation at 37 °C, which is nearly four times higher than the expected value of 117 kDa. In fact, this elevated average molecular mass value should only be taken as a lower limit; larger self-associated complexes may have been spun to the bottom of the centrifuge cell even at the lowest operating speed and thus excluded from the “window” of observation. The concentration scans of Col108 under sedimentation equilibrium in 5 mM HAc and in TES buffer at the same loading concentration and rotor speed have distinctively different shapes, further indicating the presence of higher molecular mass species in the TES solution (Fig. 2D). Because of the complex composition of a sample undergoing a progressive self-association without discrete steps of oligomerization, it is nearly impossible to fit these curves for any reliable quantitative estimate (32). Nevertheless, the AUC data unambiguously support an in-solution self-association reaction of Col108 during fibrillogenesis.

The complex assembly is reversible, and the complexes are devoid of any covalent cross-linking such as disulfide bonds. The Cys knots engineered at the ends of the triple helical domain predominantly form interchain disulfide bonds within a triple helix and cause Col108 to migrate as a trimer on SDS-PAGE under nonreducing conditions (Fig. 2C, lanes 1, 6, and 7). Sequence analysis using MSMS sequencing confirmed the purported trimer band consists of Col108 without contamination of other proteins. A secondary band just below the major trimer band is likely caused by a subpopulation of the trimer having nonoxidized -SH groups at the N or C terminus. This population is expected to acquire an unfolded conformation having a different compactness and thus migrates differently from the fully cross-linked trimers. This band is significantly decreased upon incubation in pH 7 buffer at 37 °C (Fig. 2C, lanes 6 and 7). The oxidation of -SH groups is most effective at the neutral pH. Thus, more interchain disulfide bonds can form at pH 7, and more molecules in the subpopulation become fully cross-linked triple helices. Irrespective of the incubation condition, the addition of reducing agents effectively reduces both trimer bands and yields a single, strong monomer band (Fig. 2C, lanes 2, 4, and 5). Most significantly, the data in Fig. 2C indicate that the nonoxidized -SH groups do not induce any appreciable cross-linking of the triple helices in the self-associated complexes; the aggregates can be reverted to trimers by SDS and boiling in the...
absence of reducing agents (Fig. 2C, lanes 6 and 7). The reversibility of the self-assembly under the denaturing condition is nearly complete; there are no species with molecular mass higher than the trimer present in the separation gel or in the stacking gel.

**The 35-nm Periodicity of Col108 Mini-fibrils**—The self-assembled aggregates of Col108 can be characterized as mini-fibrils with distinctive, clearly defined banding patterns by TEM (Fig. 3). The fibrillogenesis was conducted at two different temperatures (26 and 37 °C), and two incubation times were considered: 6 and 24 h. A similar banding pattern was observed for mini-fibrils formed at both temperatures and with both incubation times. The overall sizes of the mini-fibrils vary, but all have common features of tapered tips and no branches. Most mini-fibrils after 6 h of incubation are ~400–600 nm long with a diameter of ~50 nm at the center of the spindle-shaped mini-fibrils. Some mini-fibrils observed after 24 h of incubation are longer: 800 nm to over 1 µm long with the ends of individual, long mini-fibrils often being difficult to discern (Fig. 3, C and D). Still, even for long mini-fibrils, tapered tips are clearly visible, and the diameters do not grow much thicker, with most mini-fibrils having a diameter of ~75 nm and a few having a diameter of >100 nm. The triple helical conformation is a prerequisite for the periodic mini-fibril formation of Col108.

Fibrillogenesis trials using samples of Col108 that did not exhibit the typical CD spectrum of a triple helix resulted in only nonspecific aggregates without distinct banding patterns.

The cross-striated bands on the mini-fibrils indicate a specific spatial arrangement of the triple helices in the mini-fibrils. The negatively stained mini-fibrils reveal a banding pattern consisting of dark bands ~25 nm in width alternating with light bands of ~10 nm (Fig. 3, A–D). Under typical negative staining conditions, dark bands are indicative of regions where the heavy metal staining reagent can accumulate, such as the gap region of the native collagen fibrils; the light regions are considered to be stain exclusion zones, such as the overlap region of collagen fibrils. The comparable banding feature observed for Col108 led us to conclude that the Col108 mini-fibrils have alternating gap and overlap regions similar to native collagen fibrils, except the axial periodicity of Col108 is ~35 nm, including a 25-nm gap (dark band) and a 10-nm overlap zone (light band). The average value of the periodicity of mini-fibrils formed at 37 °C was determined to be 34.6 ± 1.8 nm (Fig. 3E); the standard deviation reflects the distribution of the size of the periodicity of the mini-fibrils estimated using a method with an inherent accuracy of ~3 nm (see “Experiment Procedures”). The mini-fibrils formed after incubation at 26 °C have a similar average periodicity of 34.3 ± 2.1 nm. This 35-nm axial periodicity will herein be denoted as the d-period of Col108 to highlight its comparable yet different properties from the D-period of fibrillar collagens.

Positively stained mini-fibrils also reveal a specific banding pattern along the length of the mini-fibrils (Fig. 3, F–H). The positively stained banding pattern can be characterized by three
Mini-fibril with Axial Periodicity from a Designed Peptide

FIGURE 3. The TEM pictures of Col108 mini-fibrils. A–D, the negatively stained mini-fibrils after incubation for 6 h at 37 °C (A), 24 h at 37 °C (B and C), and 6 h at 26 °C (D); The scale bars are 100 nm in A and D, 200 nm in B, and 500 nm in C. The black bars drawn next to the mini-fibrils in A, in the inset of A, and in D are 100 nm in scale. E, histogram of measured d-periodicity of mini-fibrils formed at 37 °C. F–H, the positively stained mini-fibrils after incubation for 24 h at 37 °C (F), 24 h at 26 °C (G), and 6 h at 37 °C (H). The scale bars are 100 nm in F and 200 nm in G and H. The black bars in F and in the inset of H are 100 nm in scale.

pairs of alternating thick-thin dark bands intercalated with light bands within a 110–125-nm span, a length corresponding to the size of one Col108 triple helix. Different from negative staining, positive staining is based on the direct binding of heavy metal staining salts to charged residues on the surface of the fibril. The fact that the banding pattern persists through the entire length of a positively stained mini-fibril indicates a specific, well organized assembly of the triple helices with a nearly perfect alignment of the charged residues. Any random bundling of triple helices or arrangements of triple helices with misaligned charge zones would result in an undefined distribution of the charged regions on the surface of the mini-fibril and, thus, a nonspecific staining appearance.

The axial periodicity of the Col108 mini-fibrils observed under TEM is nearly quantitatively emulated by AFM. The scanning image and resulting contour profile of a sample incubated at 37 °C for 24 h are shown in Fig. 4. The mini-fibrils appear to have the same size and shape as those seen under TEM, being ~800–1000 nm in length and 50–75 nm in diameter, and display clear cross-striations on the surfaces. The visible, bulky appearances on the surface of the mini-fibrils are due to salt crystal buildup resulting from the drying process. The scanning profiles of selected mini-fibrils revealed a periodicity of 38.0 ± 7.3 nm based on the measurements of 40 mini-fibrils (Fig. 4C). The standard deviation is more than three times greater than that estimated from TEM images because of the low resolution of the AFM data. The boundaries of the gap and overlap regions are difficult to discern because of salt deposits on the surface, and the size of the banding structure is close to the detection limit of the tapping tip used for the study. Nonetheless, the prominent periodicity is evident.

The Favorable Interactions of the 123-Residue Staggered Assembly of Col108—The axially repeating features of the Col108 mini-fibrils imply a unique, specifically ordered staggering arrangement of triple helices. The banding pattern of the mini-fibrils must be stabilized by maximized interactions between interacting residues arranged periodically along the helix. Although the amino acid sequence of Col108 has three pseudo-repeating sequence units, how this sequence periodicity relates to the axial periodicity of the mini-fibrils is not immediately apparent. The sequence alignment approach successfully devised by Hulmes et al. (10, 33) for the study of type I collagen fibrils was adopted to probe the underlying sequence features of Col108 in connection with the 35-nm periodicity of the mini-fibrils. In this approach, the total hydrophobic and electrostatic interactions between two adjacent triple helices were calculated as a function of chain stagger to determine whether maximal interactions occurred at specific residue offsets; the resulting offsets may implicate the unit repeat in the sequence of Col108 as determining the formation of the characteristic mini-fibrils. The foldon domain was treated as globular with a diameter of 2.5 nm (the diameter of the triple helix is ~1.5 nm) and as having an average of three negative charges on its surface based on the crystal structure (24).

The periodic nature of the calculated interaction graphs for helices in a parallel alignment is unmistakable (Fig. 5A). The set of peaks with shift values at multiples of 123 stands out as the staggering arrangement with optimal interactions; the peaks are especially pronounced in the total and hydrophobic interaction graphs. Because of the prominent presence of Pro residues in Col108, Pro was included in one of the calculations examining the hydrophobic interactions despite its weak hydrophobic side chain. The inclusion of Pro did not affect the periodic features of the outcome but only increased the overall magnitude of the interaction because of the increased number of residues included. The set of peaks with shifts of $n \times 123$ is observed in calculations with and without the inclusion of Pro (Fig. 5B).

Favorable interactions also exist at shift values at approximately $n \times 24$ but at much lower magnitudes. A close inspection of the sequence of Col108 indicates that clusters of charged and hydrophobic residues are in approximate alignment when two helices are arranged with shift values of $n \times 24$; such alignment is nearly completely “out of phase” when the shift values
Mini-fibril with Axial Periodicity from a Designed Peptide

are odd multiples of 12 (1 × 12, 3 × 12, 5 × 12, etc.). The electrostatic interactions are particularly reflective of the favorable interactions at this smaller staggering value, although the favorable interactions at n × 123 shifting values are also evident. It should be noted that the scale for the magnitude of the interaction is an arbitrary one. The magnitude of the peaks represents the potential of interaction and not the absolute value of the stabilizing energy. The relatively lower magnitude of the electrostatic interactions can be accounted for by recognizing that there are fewer charged residues (75 residues) compared with hydrophobic ones (120 including Pro); additionally, the interactions of like charges are set to be zero and do not contribute to the interaction (10). The actual contribution of the electrostatic interactions to the stabilization of the mini-fibrils may not be less than that of the hydrophobic interactions. By comparison, the interaction graph of an antiparallel arrangement of helices is remarkably lacking any dominating peaks (Fig. 5C); no one particular chain staggering conformation dominates the antiparallel arrangement. Although not shown, similar nondescript graphs were obtained for both hydrophobic and electrostatic interactions for an antiparallel arrangement.

The dominating and isolated interaction peaks resulting from the sequence alignments suggest that a parallel arrangement with a staggering of 123 residues affords both maximal molecular interactions and uniqueness in the association pattern during the mini-fibril assembly. Interestingly, this staggering offset of 123 residues coincides with both the size of one sequence unit of Col108 and the structural details of the mini-fibrils. Among the three pseudo-repeating sequence units of Col108, U1 and U3 each have 120 residues, whereas U2 has 126. A mutual staggering of ~123 residues would bring the Col domains and the major part of the (Gly-Pro-Pro)4 inserts of the neighboring triple helices into register and thus maximize the potential interhelical interactions of like residues. Taking the average axial rise of 0.8–0.9 nm per Gly-X-Y tripeptide of the linear conformation of the triple helix, the 123 residues of a sequence unit (41 Gly-X-YXY tripeptides) would make up a section of triple helix 35–36 nm in length, which coincides with the observed axial periodicity. The fully folded Col108 (~120 nm, including the foldon domain) would comprise ~3.3 such periods. The final fractional periodicity is a necessary feature for staggered fibrils to have gap overlap zones, and the 0.3 period agrees with the observed overlap zone in size (i.e. ~10 nm/35 nm).

A Model for the Col108 Mini-fibril with a Periodicity of 35 nm: The 1-Unit Staggered Self-assembly of Col108—A self-assembly model of Col108 was proposed that reflects the features of both the negatively stained and the positively stained Col108 mini-fibrils (Fig. 6). In this model, the 35-nm d-period corresponds to a 123-residue pseudo-repeating sequence unit of Col108; the C-terminal (GPP)4 nucleation sequence and the foldon domain form the major part of the 0.3 d overhang (Fig. 1). The mini-fibrils are formed through the lateral association of the triple helices with a mutual staggering of one sequence unit at the N terminus, creating an axially repeating d-periodicity of ~35 nm having a 0.3 d light, overlap zone and a 0.7 d dark, gap zone under negatively stained TEM images (Fig. 3, A–D). Not only does the number of residues of a sequence unit correspond well in size to the observed 35-nm axial periodicity, an offset of one sequence unit would lead to a perfect alignment of the sequence units of neighboring helices. This sequence alignment would produce maximal interactions between the charged and the hydrophobic residues of the neighboring helices as reflected in the set of n × 123 peaks depicted in interaction curve (Fig. 5, A and B). Furthermore, in the unit-staggering assembly, the spatial arrangement of the charged regions of each triple helix will be preserved in the mini-fibrils. The distribution of the charged residues within one sequence unit can be roughly characterized by two high charge-density regions, with one being slightly wider in range than the other (Fig. 6A). These charged regions are expected to appear as three pairs of dark thick-thin bands, with intercalating light strips of charge
free zones every 105 nm (the approximate length of three sequence units) under the positively stained TEM images, which are consistent with those observed in mini-fibrils (Fig. 3, F–H). Thus, the model mini-fibril using a stagger of 1 sequence unit can explain both banding patterns of negatively stained mini-fibrils and the positively stained mini-fibrils.

Although the repeating sequence architecture of Col108 warrants the alignment of the sequence units in any arrangement with an offset value of \( n \) units (where \( n = 0, 1, 2, \) or 3), only the 1-unit staggering one will generate the 35-nm \( d \)-periodicity. As shown in Fig. 6 (B and C), the exclusive 2- and 3-unit staggering arrangements will produce mini-fibrils with an axial periodicity of 70 and 105 nm, respectively. These mini-fibrils will have staining patterns easily distinguishable from those shown in Fig. 3 (A–D). An exclusive 0-unit offset assembly, on the other hand, will not form staggered long fibrils but only end on end stacks with one dimension approximately the length of a Col108 monomer. Arrangements of triple helices with a mixture of different \( n \)-unit offset values would therefore produce mini-fibrils with gaps and overlaps of varied sizes arranged axially in undefined formations. TEM images of such mini-fibrils would have unrecognizable negatively stained banding patterns. Thus, the clear and persistent presence of the 35-nm \( d \)-period of the negatively stained mini-fibrils supports the 1-unit staggering as the predominant arrangement of the associating triple helices.

DISCUSSION

In summary, our data demonstrate the self-assembly of Col108 in solution to form mini-fibrils with a specific axial structure. We propose a model in which the 35-nm period of the mini-fibrils is attributed to a mutual staggering of triple helices by one sequence unit. The sequence architecture of Col108 dictates that this 1-unit staggered arrangement not only brings the hydrophobic residues and charged residues of each unit into optimal alignment with the neighboring helices and thus maximizes the stabilizing interactions but also ensures such interactions of each unit will be repeated every 123 residues, combinatorially amplifying the interactions during lateral assembly. Combined, these synergistic interactions make the 1-unit staggered arrangement the unique conformation with the highest stability. Although we are still in the early stages of defining specific interactions in conjunction to the sequence and the mini-fibril structure, it is evident that the determinant factors for the 35-nm axial repeats of Col108 are coded within the amino acid sequence.

The assembly of the mini-fibril is driven by noncovalent interactions between the triple helices, and the robust 35-nm periodicity indicates a highly specific molecular recognition mechanism. The residues of the Col domain were selected from type I collagen for their relatively high propensity for forming a triple helical conformation, yet they may also possess certain intrinsic properties of natural collagen that bring about strong molecular interactions between associating helices. The Col domain has a relatively high content of closely spaced negatively and positively charged residues. Sequences of charged residues such as Lys-Gly-Glu and Lys-Gly-Asp are likely involved in the formation of salt bridges between the three constituent chains of a triple helix (21, 27), whereas others may contribute to interactions and molecular recognition between triple helix monomers during self-assembly. The charged residues within the Col domain are often present as pairs with opposite charges. When the clusters of charged residues are aligned between neighboring triple helices, the attractive interactions of oppositely charged residues likely dominate over the repulsion of like charges. A similar distribution of paired charged residues has also been noted for native collagen (6, 33), although any attributions this distribution lends to the specific structure or functions of collagen were never explored. The periodic insertions of the (Gly-Pro-Pro)\(_4\) sequence may also play a structural role.
favoring the 1-unit staggered assembly; their specific involvement is yet to be defined.

The multiple Cys residues at the C and N termini raised the concern that the nonoxidized -SH groups may form C to N disulfide-linked aggregates of Col108. Our data indicate that the -SH groups primarily form disulfide bonds within a triple helix and not cross-links between the triple helices in the mini-fibrils. Some interhelical disulfide bonds may form from the free -SH groups in some preparations; however, such disulfide bonds are not the deterministic factor for the specific 1-unit staggered arrangement. The dissolution of the mini-fibrils to triple helices under nonreducing conditions indicates the non-consequential role of any possible cross-linking occurring during the self-assembly. Additionally, the C to N disulfide-linked aggregates would lead to mini-fibrils with a different staining pattern (Fig. 6C); we have not observed such aggregates. Interestingly, we have also not observed the end on end, or stacked, assembly of Col108. In this arrangement, all three units of neighboring triple helices would overlap laterally and lead to the formation of aggregates with a dimension close to that of the length of a single Col108 monomer, i.e. ~120 nm. The slightly bulkier size and/or the net negative charge of the foldon domain may have imposed constraints and renders the end on end stacking unfavorable.

The role of the foldon in the mini-fibrils is intriguing. The conformation of the foldon has a diameter of 2.5 nm, which is slightly larger than the diameter of ~1.5 nm of a typical triple helix, and is packed with a defined bent angle with respect to the triple helical domain in the crystal structure of GPP10-foldon (24). However, the “kink” between the two domains is likely caused by crystal packing, and the conformation is likely to be more flexible in solution. The smooth mini-fibrils of Col108 indicate that the foldon can be accommodated, to a degree, in the packing of the helix. The fact that the foldon is located adjacent to the gap region may afford some spatial flexibility of the mini-fibrils to accommodate its larger size. On the other hand, the structural constrains of the foldon may impose certain destabilizing effects and impede the growth of the mini-fibrils. The length of the mini-fibrils are limited to approximately five to eight times the end to end lengths of the Col108 triple helix, which represents a limited extent of growth compared with that of collagen fibrils. In the modeling of the 1-unit staggered mini-fibrils, the negatively charged foldon domain appears to be surrounded by two positively charged residues from neighboring triple helices, which likely affords some stabilization. Such interactions, however, are not a determining factor for the structural specificity of the 35-nm axial periodicity, because there exist multiple clusters of positively charged residues on the Col108 triple helix. It is inconceivable to consider the foldon selectively anchors on these two particular positive residues without a more profound stabilizing mechanism in place, which conceivably brings the neighboring triple helices together and positions the foldon in the vicinity of the two residues.

FIGURE 6. The model of the Col108 mini-fibrils with 35-nm periodicity. A, the 1-unit staggering assembly of Col108 and the resultant negatively stained (top) and positively stained (bottom) banding patterns of the mini-fibrils; the dark bars indicate stained area under TEM. The three sequence units of Col108 are shown in different shades of gray. The black vertical lines in each unit show the locations of charged residues. The oval represents the foldon domain. A rod having three thick-thin charged zones indicated in black is superimposed on select Col108 helices to highlight the alignment of the charged residues. The solid double-headed arrow marks one 35-nm d-period on the negatively stained mini-fibril. The dotted-lined double-headed arrow marks the length of three sequence units (~105 nm) on the positively stained mini-fibril. B, the model of 2-unit staggering mini-fibrils. The predicted negatively stained pattern of the mini-fibrils is shown on top with the dark bar showing the stained gap region. The double-headed arrow marks one 70-nm d-period predicted for this arrangement. C, a model of mini-fibrils of 3-unit staggering arrangement. The double-headed arrow marks one 105-nm d-period predicted for this arrangement. This 3-unit arrangement can also represent the mini-fibrils driven by the interhelical S-S cross-linking. The curved lines mark the locations of clusters of Cys residues.
Mini-fibril with Axial Periodicity from a Designed Peptide

Distinct from native fibrillar collagens, Col108 contains no Hyp residues, nor does it have any telopeptides, two factors that have been under constant focus for their purported roles in the fibrillogenesis of native collagen. Despite its anticipated significance in collagen function and structure, the role of Hyp in fibrillogenesis has never been clearly defined. It was first proposed in the 1950s by Gustavson (34) that Hyp may contribute to triple helix association through either a hydration network or direct intermolecular hydrogen-bonding interactions between triple helices. The differences in crystal packing of triple helical peptide (Pro-Pro-Gly)\(_{10}\), without Hyp residues and peptides having a high Hyp content were taken as evidence that Hyp is involved in the organization of the water network surrounding the triple helix (14). It was further suggested that the formation of higher molecular aggregates of the (Pro-Hyp-Gly)\(_{10}\) peptide is mediated by hydrogen-bonded hydration networks involving Hyp (13). The concept of this hydration network, or more precisely the lack of it, was used to explain the deficiency of fibril assembly of a recombinant collagen without Hyp, from transgenic plants, under physiological conditions. Nonetheless, this unhydroxylated collagen was found to retain the ability to form fibrils with a 67-nm banding pattern, albeit only in buffers of low ionic strength (14). Another recombinant type I collagen from a yeast expression system, the rho collagen, was also reported to form “normal fibrils in vitro” despite the absence of the canonical hydroxylation pattern of Pro residues (35). The Hyp content of rho collagen is only 50% of that of the type I collagen, and the locations of the Hyp are not specified. Although the involvement of Hyp in a hydration network or a hydrogen-bonding network may contribute to the stabilization of fibrils, it is not an intrinsic factor for structural specificity or formation of collagen fibrils. Factors affected by the interactions involving Hyp are likely to include the kinetics and the optimal conditions of the fibrillogenesis, or even the diameter of the fibrils (36). Still, the absence of Hyp does not abolish the ability of a triple helix to form a periodic assembly. The ability of Col108 to self-assemble into a highly ordered, axially periodic structure is consistent with the conclusion that Hyp is not a requirement for D-staggered packing during the fibrillogenesis of collagen.

The role of telopeptides in fibrillogenesis is another debated issue. The involvement of telopeptides in the early stages of fibrillogenesis was extrapolated to involve the binding of one telopeptide to the triple helical region of a nearby collagen monomer to place “the monomers in quarter D-period stagger” and to promote the formation of a structural nucleus essential for further growth of the fibrils (11). Removal of the telopeptides by proteolytic treatment of collagens isolated from tissues was found to affect the kinetics (12, 37) and change the morphology of the fibrils (38–40) or even prevent the formation of fibrils having a banding pattern (11). Other studies, however, attributed the negative impacts on fibril formation to accidental removal of portions of the triple helix domain itself during proteolytic removal of the telopeptides. Subsequently, collagen was found to be capable of forming fibrils even after complete removal of the telopeptides (2). This latter finding challenges the notion that the telopeptides are part of the code necessary for recognition during fibrillogenesis. Although Col108 does not have the N- or C-telopeptide of collagen, it does contain a foldon domain at the C terminus. Further study is underway to elucidate any possible anchoring roles of this noncollagenous domain of Col108 during the self-assembly of the mini-fibrils.

In comparison with other nonspecific self-assemblies of triple helical peptides, the larger size and the rich collection of charged and hydrophobic residues of Col108, in addition to the repeating units of the sequences, may promote the structural specificity of the Col108 mini-fibrils. The need for more diverse sequences than simple repeats of (Pro-Pro-Gly) or (Pro-Hyp-Gly) triads for the presence of structural features among self-assembled triple helices has long been recognized (13); the effects of long range sequence repeats beyond the (Gly-X-Y) triplets, on the other hand, is much less explored. How to build such repeating units of sequences having a high variety of different residues into short synthetic peptides is challenging. Inclusion of multiple (Gly-Pro-Hyp) triplets is often necessary for the stability of short triple helical peptides (41). The larger size of Col108 may provide more extensive contacts between the associating helices. Just like protein folding in general, a critical amount of stabilizing interactions is often required for a specific structure to be selected over other competing forms, and this leads to its dominance in the free energy-driven, spontaneous process. We do not know whether the 120 residues of one sequence unit represents, or is close to, the minimum size necessary to support an axially repeating structure during assembly or whether the number three represents a “golden number” for the number of repeating sequence units. The relatively simple and yet effective Col108 system offers the opportunity to address these and an array of other fundamental questions about the fibrillogenesis of collagen and about protein design.

Acknowledgments—We thank Dr. Stephen Doty for help on TEM and Dr. James Haven for his help with the Perl script programming for the calculations of the interaction curve.

REFERENCES

1. Kadler, K. E., Hojima, Y., and Prockop, D. J. (1987) Assembly of collagen fibrils de novo by cleavage of the type I pro-collagen with procollagen C-proteinase. Assay of critical concentration demonstrates that collagen self-assembly is a classical example of an entropy-driven process. J. Biol. Chem. 262, 15696–15701

2. Kuznetsova, N., and Leikin, S. (1999) Does the triple helical domain of type I collagen encode molecular recognition and fiber assembly while telopeptides serve as catalytic domains? Effect of proteolytic cleavage on fibrillogenesis and on collagen-collagen interaction in fibers. J. Biol. Chem. 274, 36083–36088

3. Kadler, K. E., Holmes, D. F., Trotter, J. A., and Chapman, J. A. (1996) Collagen fibril formation. Biochem. J. 316, 1–11

4. Kiely, C. M., Hopkinson, I., and Grant, M. E. (1993) The collagen family: structure, assembly, and organization in the extracellular matrix. In Connective Tissue and Its Heritable Disorders (Royce, P. M., and Steinmann, B., eds) pp. 103–147, Wiley-Liss, New York

5. Prockop, D. J., and Fertala, A. (1998) The collagen fibril: the almost crystalline structure. J. Struct. Biol. 122, 111–118

6. Piez, K. A., and Miller, A. (1974) The structure of collagen fibrils. J. Supramol. Struct. 2, 121–137

7. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. Science 266, 75–81
8. Hodge, A. J., and Petruska, J. A. (1963) Aspects of Protein Structure (Ramachandran, G. N., ed) p. 289, Academic Press, New York

9. Doyle, B. B., Hulmes, D. J., Miller, A., Parry, D. A., Piez, K. A., and Woodhead-Galloway, J. (2007) Axially projected collagen structures. Proc. R. Soc. Lond. B Biol. Sci. 274, 37–46

10. Hulmes, D. J., Miller, A., Parry, D. A., Piez, K. A., and Woodhead-Galloway, J. (1973) Analysis of the primary structure of collagen for the origins of molecular packing. J. Mol. Biol. 79, 137–148

11. Prockop, D. J., and Fertala, A. (1998) Inhibition of the self-assembly of collagen I into fibrils with synthetic peptides. Demonstration that assembly is driven by specific binding sites on the monomers. J. Biol. Chem. 273, 15598–15604

12. Comper, W. D., and Veis, A. (1977) The mechanism of nucleation for in vitro collagen fibril formation. Biopolymers 16, 2113–2131

13. Kar, K., Amin, P., Bryan, M. A., Persikov, A. V., Mohns, A., Wang, Y. H., and Brodsky, B. (2006) Self-association of collagen triple helic peptides into higher order structures. J. Biol. Chem. 281, 33283–33290

14. Perret, S., Merle, C., Bernocco, S., Berland, P., Garrone, R., Hulmes, D. J., Theisen, M., and Ruggiero, F. (2001) Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation. J. Biol. Chem. 276, 43693–43698

15. Kotch, F. W., and Raines, R. T. (2006) Self-assembly of synthetic collagen triple helices. Proc. Natl. Acad. Sci. U.S.A. 103, 3028–3033

16. Paramonov, S. E., Jun, H. W., and Hartgerink, J. D. (2006) Self-assembly of peptide-aminophila nanofibers: the roles of hydrogen bonding and amphiphilic packing. J. Am. Chem. Soc. 128, 7291–7298

17. Paramonov, S. E., Jun, H. W., and Hartgerink, J. D. (2006) Modulation of peptide-aminophila nanofibers via phospholipid inclusions. Biomacromolecules 7, 24–26

18. Cjeas, M. A., Kinney, W. A., Chen, C., Leo, G. C., Toung, B. A., Vinter, J. G., Joshi, P. P., and Maryanoff, B. E. (2007) Collagen-related peptides: self-assembly of short, single strands into a functional biomaterial of micrometer scale. J. Am. Chem. Soc. 129, 2202–2203

19. Kar, K., Ibrar, S., Nanda, V., Getz, T. M., Kunapoli, S. P., and Brodsky, B. (2009) Aromatic interactions promote self-association of collagen triple-helical peptides to higher-order structures. Biochemistry 48, 7959–7968

20. Rele, S., Song, Y., Apkarian, R. P., Qu, Z., Conticello, V. P., and Chaikoff, E. L. (2007) D-periodic collagen-mimetic microfibers. J. Am. Chem. Soc. 129, 14780–14787

21. Xu, K., Nowak, I., Kirchner, M., and Xu, Y. (2008) Recombinant collagen studies link the severe conformational changes induced by osteogenesis imperfecta mutations to the disruption of a set of interchain salt bridges. J. Biol. Chem. 283, 34337–34344

22. Frank, S., Kammerer, R. A., Mechling, D., Schultz, T., Landwehr, R., Bann, J., Gao, Y., Lustig, A., Bächinger, H. P., and Engel, J. (2001) Stabilization of short collagen-like triple helices by protein engineering. J. Mol. Biol. 308, 1081–1089

23. Horcas, I., Fernández, R., Gómez-Rodríguez, J. M., Colchero, J., Gómez-Herrero, J., and Baro, A. M. (2007) WSXM: a software for scanning probe microscopy and a tool for nanotechnology. Rev. Sci. Instrum. 78, 013705

24. Stetefeld, J., Frank, S., Jenifer, M., Schultz, T., Kammerer, R. A., Boudko, S., Landwehr, R., Okuyama, K., and Engel, J. (2003) Collagen stabilization at atomic level: crystal structure of designed (GlyProPro)10Foldon. Structure 11, 339–346

25. Bächinger, H. P., Bruckner, P., Timpl, R., Prockop, D. J., and Engel, J. (1980) Folding mechanism of the triple helix in type-III collagen and type-III pN-collagen. Eur. J. Biochem. 106, 619–632

26. Boudko, S., Frank, S., Kammerer, R. A., Stetefeld, J., Schultz, T., Landwehr, R., Lustig, A., Bächinger, H. P., and Engel, J. (2002) Nucleation and propagation of the collagen triple helix in single-chain and trimerized peptides: transition from third to first order kinetics. J. Mol. Biol. 317, 459–470

27. Persikov, A. V., Ramshaw, J. A., and Brodsky, B. (2005) Prediction of collagen stability from amino acid sequence. J. Biol. Chem. 280, 19343–19349

28. Mechnling, D. E., and Bächinger, H. P. (2000) The collagen-like peptide (GER)15GPPCG forms pH-dependent covalently linked triple helical trimers. J. Biol. Chem. 275, 14532–14536

29. Feng, Y., Melacini, G., Taulane, J. P., and Goodman, M. (1996) Acetyl-terminated and template-assembled collagen-based polypeptides composed of Gly-Pro-Hyp sequences: 2. synthesis and conformational analysis by circular dichroism, ultraviolet absorbance and optical rotation. J. Am. Chem. Soc. 118, 10351–10358

30. Ottl, J., Gabriel, D., Murphy, G., Knäuper, V., Tominaaga, Y., Nagase, H., Kröger, M., Tschesche, H., Bode, W., and Moroder, L. (2000) Recognition and catabolism of synthetic heterotrimeric collagen peptides by matrix metalloproteinases. Chem. Biol. 7, 119–132

31. Heidemann, E., and Roth, W. (1982) Synthesis and investigation of collagen model peptides. Adv. Polymer Sci. 43, 143–203

32. Xu, Y. (2004) Characterization of heterogeneity of self-associating systems using equilibrium sedimentation techniques. Biophys. Chem. 108, 141–163

33. Hulmes, D. J., Miller, A., Parry, D. A., and Woodhead-Galloway, J. (1977) Fundamental periodicities in the amino acid sequence of the collagen α1 chain. Biochem. Biophys. Res. Commun. 77, 574–580

34. Gustavson, K. H. (1955) The function of hydroxyproline in collagens. Nature 175, 70–74

35. Olsen, D. R., Leigh, S. D., Chang, R., McMullin, H., Ong, W., Tai, E., Chisholm, G., Birk, D. E., Berg, R. A., Hitzeman, R. A., and Toman, P. D. (2001) Production of human type I collagen in yeast reveals unexpected new insights into the molecular assembly of collagen trimers. J. Biol. Chem. 276, 24038–24043

36. Prockop, D. J., and Kivirikko, K. I. (1969) Effect of polymer size on the inhibition of procollagen proline hydroxylase by polyproline II. J. Biol. Chem. 244, 4838–4842

37. Helseth, D. L., Jr., and Veis, A. (1981) Collagen self-assembly in vitro: differentiating specific telopeptide-dependent interactions using selective enzyme modification and the addition of free amino telopeptide. J. Biol. Chem. 256, 7118–7128

38. Ghosh, S. K., and Mitra, H. P. (1975) Oblique banding pattern in collagen fibrils reconstituted in vitro after trypsin treatment. Biochim. Biophys. Acta 405, 340–346

39. Drake, M. P., Davison, P. F., Bump, S., and Schmitt, F. O. (1966) Action of proteolytic enzymes on tropocollagen and insoluble collagen. Biochim. Biophys. Acta 113, 301–312

40. Leibovich, S. J., and Weiss, J. B. (1970) Electron microscope studies of the effects of endo- and exopeptidase digestion on tropocollagen. A novel concept of the role of terminal regions in fibrillogenesis. Biochim. Biophys. Acta 214, 445–454

41. Brodsky, B., and Ramshaw, J. A. (1997) The collagen triple helix structure. Matrix Biol. 15, 545–554