Actinidain-hydrolyzed Type I Collagen Reveals a Crucial Amino Acid Sequence in Fibril Formation*

Saori Kunii‡, Koichi Morimoto‡,‡, Kouhei Nagai‡, Takuya Saito‡, Kenji Sato‡, and Ben’ichiro Tonomura‡

From the §Department of Biotechnological Science, Kinki University, 930 Nishimitani, Kinokawa, Wakayama 649-6493, the §Graduate School of Medicine, Saint Marianna University School of Medicine, 2-16-1 Sugao, Kawasaki, Kanagawa 216-8511, and the ¶Division of Applied Life Sciences, Graduate School of Life and Environmental Sciences, Kyoto Prefectural University, 1-5 Shimogamonakaragichou, Kyoto, Kyoto 606-8522, Japan

We investigated the ability of type I collagen telopeptides to bind neighboring collagen molecules, which is thought to be the initial event in fibrillogenesis. Limited hydrolysis by actinidain protease produced monomeric collagen, which consisted almost entirely of protease produced monomeric collagen, which consisted of actinidain-hydrolyzed collagen exhibited unique self-assembly, as if at an intermediate stage, and formed a novel suprastructure characterized by poor fibrillogenesis. Then, the N- and C-terminal sequences of chicken type I collagen hydrolyzed by actinidain or pepsin were determined by Edman degradation and de novo sequence analysis with matrix-assisted laser desorption ionization tandem mass spectrometry, respectively. In the C-te- lopeptide region of the α1 chain, pepsin cleaved between Asp¹⁰³⁵ and Phe¹⁰³⁶, and actinidain between Gly¹⁰³² and Gly¹⁰³³. Thus, actinidain-hydrolyzed α1 chain is shorter at the C terminus by three residues, Gly¹⁰³³, Phe¹⁰³⁴, and Asp¹⁰³⁵. In the α2 chain, both proteases cleaved between Glu¹⁰⁵⁰ and Val¹⁰⁵¹. We demonstrated that a synthetic nonapeptide mimicking the α1 C-terminal sequence including GFD weakly inhibited the self-assembly of pepsin-hydrolyzed collagen, whereas it remarkably accelerated that of actinidain-hydrolyzed collagen. We conclude that the specific GFD sequence of the C-te-lopeptide of the α1 chain plays a crucial role in stipulating collagen suprastructure and in subsequent fibril formation.

Type I collagen is the most abundant protein in connective tissue of all vertebrates. It forms highly ordered fibrils that are generally thought to provide mechanical strength and regulate cell function (1–5). Collagen is formed from tightly interwoven heterotrimeric chains of α1 chains and one α2 chain in a triple-helical coiled-coil structure. The triple helix consisting of Gly-Xaa-Yaa repeats is ~1,000 amino acid residues in length and is resistant to proteolysis except for specific collagenases that can cleave the helix (6, 7). The high content of 4-hydroxyproline in the Yaa position stabilizes the triple-helical structure (8, 9).

Observations of in vitro collagen fibril formation suggest that the collagen molecule has sufficient structural information to assemble spontaneously under suitable conditions. This structural information has been partially revealed by atomic force microscopy, electron microscopy, spectroscopic analysis, and x-ray diffraction (10–15). From this information, it appears that the telopeptide regions play an important role in the rate of fibril formation and in azimuthal and lateral growth of fibrils. Indeed, fibril formation of pepsin-hydrolyzed collagen (PHCol) that has been partially hydrolyzed at the N- and C-te-lopeptide domains progresses more slowly than that of acid-soluble collagen (ASCol) (16, 17). The recognition and association of collagen N- and C-telopeptides have been partly demonstrated in vitro (18–25). The results imply that the N- and C-telopeptide regions are required to “seed” the interaction between collagen molecules. It was also concluded that the triple-helical structure is essential for fibril formation (4, 26).

Our experimental approach to understand the mechanism of fibril formation was to investigate type I collagen that had been partially hydrolyzed by actinidain protease (EC 3.4.22.14). Actinidain is a cysteine protease found in kiwi fruit (Actinidia delicosa) and is a member of the papain superfamily with a broad specificity toward a variety of substrates (27–29). We previously found that actinidain-hydrolyzed collagen (AHCol) of tuna retained its triple-helical structure at pH 4.0 and showed the same thermal stability as that of ASCol, as judged by circular dichroism (CD) spectroscopy. AHCol, however, was found to be essentially monomeric with no cross-linkages, and therefore it provides a unique system for further characterization (30, 31).

Knowledge of the exact amino acid sequence is vital for understanding the behavior of collagen, including fibril formation. Although the N-terminal amino acid sequence of collagen can be determined by Edman degradation, there has been no reliable method for determining the C-terminal sequence. Using tandem mass spectrometry (MS/MS) and known amino acid sequence information, we recently determined the amino acid sequences of novel proteins following trypsin digestion (32, 33). We employed this method in the current work to iden-

* This work was supported in part by Grant-in-aid for Scientific Research 16500307 (to K. M.), a grant-in-aid for the 21st Century COE Program “Centers of Excellence” of the Ministry of Education, Culture, Sports, Science and Technology of Japan, and by the Wakayama Prefecture Collaboration of Regional Entities for the Advancement of Technological Excellence of the Japan Science and Technology Agency.

‡ The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. 1–5 and Table 1.

‡† To whom correspondence should be addressed. Tel.: 81-736-77-3888; Fax: 81-736-77-4734; E-mail: morimoto@waka.kindai.ac.jp.

2 The abbreviations used are: PHCol, pepsin-hydrolyzed collagen; ASCol, acid-soluble collagen; AHCol, actinidain-hydrolyzed collagen; MS, mass spectrometry; MS/MS, tandem mass spectrometry; SEM, scanning electron microscopy; MMP-1, matrix metalloproteinase-1; MALDI-TOF/TOF, matrix-assisted laser desorption ionization tandem mass spectrometry.
Crucial Amino Acid Sequence for Collagen Fibrillogenesis

identify the C-terminal amino acid sequence of collagen preparations. In this study, by comparing the properties of ASCöl, PHCol, and AHCol derived from chicken skin, we identified an amino acid sequence that is required for the fibril formation of type I collagen.

EXPERIMENTAL PROCEDURES

Preparation of ASCöl, PHCol, and AHCol from Chicken Skin—ASCöl and PHCol were prepared from chicken skin according to the method of Morimoto et al. (30). Actinidain was isolated from kiwi fruit as described (29). Crude ASCöl in 20 mM sodium acetate buffer, pH 4.0 (final concentration 2 mg/ml), was incubated with 0.15% actinidain (w/v) at 20 °C for 7 days. AHCol was salted out by adding NaCl to a final concentration of 2.5 M. The resulting PHCol and AHCol samples were dialyzed against ultrapure water to remove small molecule contaminants and then lyophilized. The purity of the ASCöl, PHCol, and AHCol samples was assessed by 5% SDS-PAGE in the presence of 4 M urea (30). Protein molecular mass markers were obtained from APRO Life Science Institute Inc. (Naruto, Japan). After electrophoresis, proteins were stained with Coomassie Brilliant Blue R-250. The resulting PHCol and AHCol samples were digested with trypsin, then lyophilized. The purity of the ASCöl, PHCol, and AHCol samples was confirmed by CD spectra as reported (30, 31). The far-UV CD spectrum of the collagen preparations was measured by scanning through the range of 200–250 nm with a CD spectropolarimeter J-820 (JASCO, Tokyo, Japan). Measurement was carried out using a 1.0 mg/ml solution of ASCöl, PHCol, or AHCol in 20 mM phosphoric acid (pH 2.2) in a quartz cell with a 0.1-cm optical path length at 20 °C.

Morphological Observation of Self-assembly of PHCol and AHCol Using Scanning Electron Microscopy (SEM)—PHCol and AHCol (final concentration 3 mg/ml) were dissolved in 15 mM acetic acid for 20 h at 4 °C. Four glass disks (13 mm) were coated with the PHCol or AHCol solution and air-dried at 20 °C. Then, the glass disks were placed in 50 mM sodium phosphate buffer, pH 7.4, at 37 °C to start self-assembly. After an incubation period of 10, 30, 90, or 120 min, self-assembly was stopped by placing each glass disk in 2% glutaraldehyde at room temperature for 1 h. Sample preparations for SEM observation were carried out according to the methods of Nakayama et al. (34). Collagen self-assembly was observed under an S-900 ultrahigh resolution SEM (Hitachi) with an accelerating voltage of 10 kV.

N-terminal Amino Acid Sequence Analysis of PHCol and AHCol—The α1 and α2 chains of PHCol and AHCol were separated by 5% SDS-PAGE and electrophorized on a polyvinylidene fluoride membrane at 180 mA for 1 h. Each protein spot corresponding to the α1 or α2 chain was used to determine the N-terminal amino acid sequence using a Procise 494 HT Protein Sequencing system (Applied Biosystems, Inc., Carlsbad, CA).

Matrix Metalloproteinase-1 (MMP-1) Hydrolysis of PHCol and AHCol—To determine the C-terminal amino acid sequence of PHCol and AHCol using MS/MS analysis, samples were first hydrolyzed with collagenase to obtain shorter polypeptide chains according to the published methods (7, 35). MMP-1-mediated hydrolysis was performed by the addition of 160 μl of 0.5 unit/ml MMP-1 solution (derived from active human fibroblasts; EC 3.4.24.7; Life Laboratory Corporation, Yamagata, Japan) to 200 μl of a 1.0 mg/ml PHCol (or AHCol) solution and then incubated at 37 °C for 10 days. The resulting fragments of the α1 and α2 chains were separated by 11% SDS-PAGE and stained with Coomassie Brilliant Blue R-250. The larger 3/4 fragment and the smaller 1/4 fragment generated from PHCol or AHCol by MMP-1 were named TCα1 and TCα2, respectively.

C-terminal Amino Acid Sequence Analysis of PHCol and AHCol with MALDI-MS/MS—Each polypeptide corresponding to the TCα1 and TCα2 chains was digested and purified, and the resulting peptides were analyzed using a 4700 MALDI-TOF/TOF MS (Applied Biosystems Inc.) according to our published methods (32, 33). MS/MS spectra were measured in the collision-induced dissociation-off mode. Prediction of the amino acid sequence from MS/MS data was performed using DeNovo Explorer™ (version 1.22; Applied Biosystems Inc.) with an MS/MS tolerance of 0.2 Da.

Turbidity Measurements of PHCol and AHCol in the Presence of Five Synthetic Peptides—The “cold start” procedure for fibril formation was used to observe collagen self-assembly (16, 36, 37). The self-assembly of collagen molecules was detected by monitoring turbidity as observed by an increase in optical density at 313 nm. At 20 °C, a 2.0 mg/ml collagen solution in 10 mM sodium acetate buffer, pH 4.0, was diluted 10-fold with 50 mM sodium phosphate buffer, pH 7.4; the final collagen concentration in the neutralized solution was 0.20 mg/ml (0.67 μM). The neutralized solution was poured into a cuvette with a 1-cm path length at 20 °C, and then heated to 37 °C. The turbidity of the solution was measured at 1-min intervals using a UV-2200A spectrophotometer (Shimadzu, Kyoto, Japan). Five synthetic nonapeptides mimicking the sequence around the C terminus of PHCol and AHCol were purchased from Toray Research Center Inc. (Tokyo, Japan). Each individual nonapeptide (PGPPGSG, PGPPSGGF, PSSGFDFSF, GFDFSFLPQ, and FSFLPQPPQ) was added singly to a neutralized solution of collagen at up to a 400-fold molar excess over collagen, the molecular weight of which was assumed to be 300,000. The turbidity change of the mixture was measured as above. The dose dependence of PSSGFDFSF at 10-, 50-, 100-, or 400-fold molar excess over PHCol or AHCol was examined. The following characterizing parameters were estimated manually from the turbidity progress curve (16, 36, 37): tlag, the time at the end of the lag phase; tmax, the time at the maximum fibril growth rate; and (dα/dt)max, the maximum fibril growth rate.

RESULTS

Preparation of ASCöl, PHCol, and AHCol—The components of ASCöl, PHCol, and AHCol were analyzed by 5% SDS-PAGE (see supplemental Fig. 1). The molecular mass and relative amount of each component are shown in Table 1. These data indicate that the hydrolysis site for collagen must differ between pepsin and actinidain.

Triple-helical Structure of ASCöl, PHCol, and AHCol as Determined by CD Spectroscopy—The CD spectrum for ASCöl was typical of a triple-helical structure, with a peak maximum at
Two N-terminal sequences, H2N-ADFGP- and H2N-DFGPG-, H2N-GPGPM-, and H2N-LMGPR-, were found in the PHCol and AHCol respectively. Scale bar, 300 nm. The spectra for AHCol and PHCol were identical for various periods.

Molecular mass and relative amount of each component

| Chain | ASCol | PHCol | AHCol |
|-------|-------|-------|-------|
| β11   | 259   | 254   | 253   |
| β12   | 243   | 241   | 240   |
| α1    | 132   | 129   | 127   |
| α2    | 116   | 113   | 112   |

Comparison of these sequences with the entire amino acid sequence of chicken type I collagen (Gallus gallus) deposited in the UniProtKB/Swiss-Prot data base (P02457 and P02467). The cleavage sites of pepsin and actinidain as determined by N-terminal sequence analysis are indicated by vertical lines.

Comparison of the sequences with the entire amino acid sequence of chicken type I collagen (Gallus gallus) deposited in the UniProtKB/Swiss-Prot data base (P02457, version 80; P02467, version 87) identified the cleavage sites of pepsin and actinidain in the N-telopeptide region (Fig. 2 and see "Discussion").

**MMP-1 Hydrolysis of PHCol and AHCol**—The SDS-PAGE result of AHCol after MMP-1 hydrolysis was similar to that of PHCol (see supplemental Fig. 3). Two fragments, TCα1α1 (96 kDa) and TCβα1 (29 kDa), were obtained from each α chain. The β2 chain was also hydrolyzed into two fragments, TCβα2 (86 kDa) and TCαα2 (26 kDa). Only the TCα fragments were subjected to further analysis, however, because they were thought to have been derived from the C terminus of PHCol and AHCol, by analogy with previous reports (7, 35).

**C-terminal Amino Acid Sequences of PHCol and AHCol as Determined by MS/MS**—The C-terminal amino acid sequence was determined by de novo sequence analysis using MALDI-TOF/TOF. Fig. 3, A and B, shows the MALDI-TOF MS spectra of the tryptic digests of TCαα1 chains of PHCol and AHCol, respectively. To identify the specific peak originating from the C-terminal peptide, the two spectra were compared, and peaks at m/z 2654.32 for PHCol and m/z 2335.11 for AHCol were identified as unique to each collagen preparation. The amino acid sequences of these peaks were further analyzed by MALDI-TOF/TOF and identified, by y- and b-series ions, to be GRTGEVGPGPOGPOGPOGPOGPPSG for the peak at m/z 2654.32 and GRTGEVGPGPOGPOGPOGPOGPPSG for the peak at m/z 2335.11 (see supplemental Fig. 4), where O refers to hydroxyproline. In both cases, four hydroxyproline residues were identified at the same positions as those in the database sequence.

The tryptic digest of the TCβα2 chain showed the same peak profile in MALDI-TOF MS for both PHCol and AHCol (Fig. 4). We identified the C-terminal fragment as the peak at m/z 2598.16, because this was the only peak that did not contain a C-terminal K or R, as determined by a preliminary MS/MS analysis. This peak was applied to MS/MS sequencing analysis, and the amino acid sequence was determined to be GSHGSQPGAPGPGPOGPOGPOGGY (see supplemental Fig. 5). Thus, the C-terminal amino acid sequences of both TCαα2 chains were similar using the same sequencing method, and four hydroxyproline residues were identified in the sequence. The putative pepsin and actinidain cleavage sites of the C-telopeptide region are shown in Fig. 5. Therefore, a clear

**TABLE 1**

Molecular mass and relative amount of each component

| Chain | ASCol kDa | PHCol Content (%) | AHCol Content (%) |
|-------|-----------|-------------------|-------------------|
| β11   | 259       | 27                | 17                |
| β12   | 243       | 26                | 12                |
| α1    | 132       | 25                | 43                |
| α2    | 116       | 22                | 28                |

**FIGURE 1.** SEM snapshots of PHCol and AHCol suprastructures after various incubation times. Solutions of PHCol and AHCol prepared in 15 mM acetic acid (unfolding conditions) were placed onto glass disks and air-dried. Self-assembly of the collagen was initiated by incubating the disks in 50 mM sodium phosphate buffer, pH 7.4, at 37 °C, and then assembly was followed for various periods. Upper and lower micrographs are of PHCol and AHCol, respectively. Scale bar, 300 nm.

**FIGURE 2.** N-terminal amino acid sequences of the α1 and α2 chains of chicken type I collagen. The registered amino acid sequence data are available in the UniProtKB/Swiss-Prot data base (P02457, version 80; P02467, version 87) identified the cleavage sites of pepsin and actinidain in the N-telopeptide region (Fig. 2 and see “Discussion”).

**FIGURE 3.** N-terminal amino acid sequences of the β1 and β2 chains of chicken type I collagen.

**FIGURE 4.** Mass spectrum of the C-terminal fragment of PHCol after hydrolysis with pepsin. The spectra for AHCol were identical to those of PHCol.

**FIGURE 5.** Mass spectrum of the C-terminal fragment of AHCol after hydrolysis with pepsin. The spectra for PHCol were identical to those of AHCol.

**Crucial Amino Acid Sequence for Collagen Fibrillogenesis**

The resulting SEM snapshots of self-assembly of PHCol and AHCol are shown in Fig. 1, illustrating the changes in collagen suprastructure at various incubation times. PHCol spontaneously formed typical fibrils, with fibrils increasing their diameter from 10 nm to 500 nm during the 120-min incubation period. On the other hand, AHCol showed a novel nanoscale meshwork suprastructure with a “spider-web” appearance. No fibril growth was observed during the incubation period.

**N-terminal Amino Acid Sequences of PHCol and AHCol Chains as Determined by Edman Degradation**—Two N-terminal sequences, H2N-YDEKS- and H2N-VAVPG-, were found in the α1 chain of PHCol, and two N-terminal sequences, H2N-VAVPG- and H2N-VPGPM-, were found in the α1 chain of AHCol. Two N-terminal sequences, H2N-ADFGP- and H2N-FGPGP-, were also found in the α2 chain of PHCol, and three sequences, H2N-DFGPG-, H2N-GPGPM-, and H2N-LMGPR-, were found in the α2 chain of AHCol. The N-terminal amino acid sequences therefore showed rather minor differences between PHCol and AHCol.

**Morphological Observation of PHCol and AHCol Using SEM**—The registered amino acid sequence data are available in the UniProtKB/Swiss-Prot data base (P02457, version 80; P02467, version 87) identified the cleavage sites of pepsin and actinidain in the N-telopeptide region (Fig. 2 and see “Discussion”).

**Discussion**

The tryptic digest of the TCBβ chain showed the same peak profile in MALDI-TOF MS for both PHCol and AHCol (Fig. 4). We identified the C-terminal fragment as the peak at m/z 2598.16, because this was the only peak that did not contain a C-terminal K or R, as determined by a preliminary MS/MS analysis. This peak was applied to MS/MS sequencing analysis, and the amino acid sequence was determined to be GSHGSQPGAPGPGPOGPOGPOGPPSGY (see supplemental Fig. 5). Thus, the C-terminal amino acid sequences of both TCBα2 chains were similar using the same sequencing method, and four hydroxyproline residues were identified in the sequence. The putative pepsin and actinidain cleavage sites of the C-telopeptide region are shown in Fig. 5. Therefore, a clear
of a nonapeptide was kinetically monitored. Five synthetic nonapeptides (described under “Experimental Procedures”) were tested. Without the addition of the nonapeptides, PHCol, which formed typical fibrils as seen by SEM (Fig. 1), showed a high rate of increase in turbidity, with $t_{\text{lag}}$ of 32 min, $t_{\text{max}}$ of 57 min, and $(dA/dt)_{\text{max}}$ of $12.3 \times 10^{-3}$ OD·min$^{-1}$. However, AHCol showed a longer lag time and a lower rate with a $t_{\text{lag}}$ of 100 min, $t_{\text{max}}$ of 155 min, and $(dA/dt)_{\text{max}}$ of $5.0 \times 10^{-3}$ OD·min$^{-1}$ (Fig. 6A).

The addition of four of the five nonapeptides did not affect the self-assembly of either PHCol or AHCol. Interestingly, however, one nonapeptide, PSGGFDFSF (400-fold molar concentration), weakly inhibited PHCol self-assembly but markedly accelerated AHCol self-assembly (Fig. 6A, c and c'). The turbidity parameters of PHCol in the presence of this nonapeptide changed to: $t_{\text{lag}}$, 35 min; $t_{\text{max}}$, 67 min; and $(dA/dt)_{\text{max}}$, 9.3 $10^{-3}$ OD·min$^{-1}$. On the other hand, those parameters for AHCol in the presence of this nonapeptide were: $t_{\text{lag}}$, 50 min; $t_{\text{max}}$, 85 min; and $(dA/dt)_{\text{max}}$, 8.7 $10^{-3}$ OD·min$^{-1}$. Moreover, the effect of the relative molar excess of PSGGFDFSF (10–400-fold) on the self-assembly of PHCol and AHCol was examined (Fig. 6, B and C). Substantial differences were observed in the values of the $t_{\text{lag}}$ and $t_{\text{max}}$ for AHCol (see supplemental Table 1).

This nonapeptide appeared to affect the initial step in AHCol self-assembly.

**DISCUSSION**

Collagen fibril formation is a multistep process occurring in a sequential manner between adjacent molecules. The intrinsic propensity of collagen to undergo fibrillogenesis must be connected to the amino acid sequence of the N- and C-terminal domains because truncation of either or both of the N- and C-terminal telopeptide regions leads to impaired collagen fibrillogenesis (19–22, 24–26). In this study, we compared the structural features of two types of protease-hydrolyzed collagen molecules, PHCol and AHCol, obtained from limited hydrolysis by pepsin and actinidain, respectively, from chicken skin-derived ASCol. These collagen preparations were chosen because PHCol and AHCol derived similarly from tuna skin show interesting structural differences (30, 31). Full sequence information of tuna skin collagen is not available. Therefore, we
chose chicken skin as the source of collagen for this study because the chicken collagen sequence is known.

Components of ASCol, PHCol, and AHCol as Determined by SDS-PAGE—From SDS-PAGE analysis (see supplemental Fig. 1), PHCol showed at least five bands, corresponding to α1, α2, β11, β12, and γ polypeptide chains, whereas AHCol exhibited mainly two bands, corresponding to chains α1 and α2. The relative content of α1 and α2 chains increased, and that of β11 and β12 chains decreased following the limited hydrolysis by pepsin or actinidain (Table 1). AHCol can be considered as an essentially monomeric collagen in which 58% of molecules are composed of two α1 chains and 32% of α2 chains, with few cross-linkages between α chains. These results are consistent with our previous data on tuna AHCol (30, 31).

Morphological Change during Self-assembly as Observed by SEM—As shown in Fig. 1, after a 10-min incubation, collagen molecules appeared to adhere to adjacent molecules at specific binding positions. No preferential lateral orientation was observed, indicating that the early stage of self-assembly was controlled by the binding between triple helices. After 30 min, the PHCol molecules formed with each other and clearly controlled by the binding between triple helices. After 30 min, binding positions. No preferential lateral orientation was observed. Morphological change in turbidity for the self-assembly of PHCol and AHCol in the presence of synthetic nonapeptides and the effect of PGSGDFDFS concentration on the turbidity change for the self-assembly of PHCol and AHCol. Self-assembly was monitored by measuring the absorbance at 313 nm. A, PHCol (0.67 μM) with PGPPGPPGS (α), PGPPG-GFD (β), PGSGDFDFS (c), GDFDFSFLPQ (d), or FSFLQPPOQ (e), and AHCol (0.67 μM) with PGPPGPPGS (α′), PGPPGSGFD (β′), PGSGDFDFS (c′), GDFDFSFLPQ (d′), or FSFLQPPOQ (e′). Each nonapeptide was added to the collagen solutions in a 400-fold molar concentration. B, PGSGDFDFS concentration of 10-fold (α), 50-fold (β), 100-fold (c), or 400-fold (d) molar ratios of PHCol (0.67 μM). C, PGSGDFDFS concentration of 10-fold (α), 50-fold (β), 100-fold (c), and 400-fold (d) molar ratios of AHCol (0.67 μM).

ordered helical structures were strongly related to the molecular packing of collagen molecules. Such a difference in AHCol structure suggests that the N and C termini must differ significantly from those of PHCol. These results prompted us to determine the amino acid sequences of these termini.

N-terminal Sequence Determination and Hydrolysis Sites for Pepsin and Actinidain—Pepsin cleaves the α1 chain at two sites, Gly5–Tyr6 and Gly12–Val13. Thus, the PHCol α1 chain has at least two components, one of which contains Lys9. Pepsin cleavage of the α2 chain at Ala7–Ala8 and Asp9–Phe10 removes Lys6 from both components. Therefore, the possibility to cross-link the α1 chain via Lys9 is partly retained in PHCol. Actinidain, on the other hand, cleaves the α1 chain at two sites (Gly12–Val13, Ala14–Val15) and the α2 chain at three sites (Ala8–Asp9, Phe10–Gly11, Gly16–Leu17). AHCol, therefore, has no remaining cross-linking residues because the α1 chain Lys9 and α2 chain Lys6 are in the leaving N-telopeptide fragments. These results are consistent with the observed lower content of β11 and β12 chain in AHCol (Table 1). The H2N-(YDEK-SAG)(VA)V- sequence of the α1 chain and the H2N-(A)(D)(F)-(GPGPMG)L- sequence of the α2 chain correspond to the differences in the N-terminal sequences of PHCol and AHCol (Fig. 2).

C-terminal Sequence Determination of PHCol and AHCol—Our results were in good agreement with the registered sequence data of the α1 chain of chicken type I collagen and revealed that pepsin and actinidain cleaved the α1 polypeptide chain at Asp1035–Phe1036 and Gly1032–Gly1031, respectively (Fig. 5). The difference in the molecular mass of the two peaks at m/z 2654.32 and m/z 2335.11 in Fig. 3 is m/z 319.21, which corresponds to the mass of the intervening tripeptide HN-GFD-CO (m/z 319.33).

The results for the α2 chain C-telopeptide showed that both pepsin and actinidain cleaved the polypeptide at Glu1030–Val1031. In addition, we identified five hydroxyproline residues in a GPP repeat sequence, at positions 1013, 1016, 1019, 1022, and 1024. There are no annotation data related to hydroxyproline residues in the sequence of the chicken type I collagen α2 chain, according to the UniProt entry (http://www.uniprot.org/uniprot/P02467). We demonstrated here that both PHCol and AHCol have four GPO repeat sequences in each α1 and α2 chain (see supplemental Figs. 4 and 5). These are likely to be essential for the stability of chicken collagen.

Turbidity Measurement for the Self-assembly of PHCol and AHCol in the Presence of Mimicking Nonapeptides—The result for PHCol (Fig. 6A) is in qualitative agreement with published data (16, 36, 37). The tlag of AHCol, however, was much longer than that of PHCol, and the fibril growth rate decreased to ~41% of that of PHCol. Despite the fact that the amino acid sequence of AHCol is generally similar to that of PHCol, the kinetics of self-assembly differ markedly between these collagens. It has been suggested that the C-terminal sequence of collagen plays a substantive role in fibril formation (20, 22, 24, 25). We have demonstrated that the AHCol α1 chain has a unique C terminus (Fig. 5). It is therefore worthwhile to investigate AHCol self-assembly further. Interestingly, we found that only one nonapeptide, PGSGDFDFS, among five that mimic the C-telopeptide region of the α1 chain, influenced the turbidity
kinetics. The N-terminal six residues of PSGGFDFSF overlap the C-terminal sequence of PHCol and thus may compete with PHCol for the binding site on adjacent collagen molecules (Fig. 6B). On the other hand, for AHCol, in which the sequence GFD is missing and only three residues at the C terminus overlap the N-terminal sequence of the nonapeptide, the nonapeptide seems to behave as if a piece of mending tape by filling the vacant space made by truncation (Fig. 6C). The seemingly opposite effects of this nonapeptide on self-assembly may provide an important clue to understanding the initial step of fibrillogenesis, although the molecular mechanism remains to be elucidated.

This study has directly shown that the primary structure of the N and C termini of chicken type I collagen influences the mechanism of fibrillogenesis and the resulting suprastructure and that the GFD sequence of the \(\alpha_1\) chain is required for lateral growth of fibrils. Interestingly, the GFD sequence in the C-terminal hypothesis that the presence of the collagen. In conclusion, several lines of evidence support our hypothesis that the presence of the \(\alpha_1\) chain telopeptide region containing a GFD or GYD sequence is especially important for conferring orientation specificity. PHCol and AHCol from chicken type I collagen, for which the N- and C-terminal amino acid sequences have been determined, are excellent choices for the study of self-assembly and suprastructure formation.

REFERENCES

1. Veis, A., and George, A. (1994) Extracellular Matrix Assembly and Structure (Yurchenco, P. D., Birk, D. E., and Mecham, R. P., Mecham, eds) 1st Ed., pp. 15–45, Academic Press, Inc., San Diego, CA
2. Kadler, K. E., Holmes, D. F., Trotter, J. A., and Chapman, J. A. (1996) Biochem. J. 316, 1–11
3. Di Lullo, G. A., Sweeney, S. M., Körkkö, J., Ala-Kokko, L., and San Antonio, J. D. (2002) J. Biol. Chem. 277, 4223–4231
4. Sweeney, S. M., Orgel, J. P., Fertala, A., McAuliffe, J. D., Turner, K. R., Di Lullo, G. A., Chen, S., Antipova, O., Perumal, S., Ala-Kokko, L., Forlino, A., Cabral, W. A., Barnes, A. M., Marini, J. C., and San Antonio, J. D. (2008) J. Biol. Chem. 283, 21187–21197
5. Chung, H. J., Steplewski, A., Chung, K. Y., Uitto, J., and Fertala, A. (2008) J. Biol. Chem. 283, 25879–25886
6. Perumal, S., Antipova, O., and Orgel, J. P. R. O. (2008) Proc. Natl. Acad. Sci. U.S.A. 105, 2824–2829
7. Chung, L., Dinakarpandian, D., Yoshida, N., Lauer-Fields, J. L., Fields, G. B., Visse, R., and Nagase, H. (2004) J. EMBO J. 23, 3020–3030
8. Bann, I. G., and Bächinger, H. P. (2000) J. Biol. Chem. 275, 24466–24469
9. Slater, D. A., Miles, C. A., and Bailey, A. J. (2003) J. Mol. Biol. 329, 175–183
10. Paige, M. F., Rainey, J. K., and Goh, M. C. (1998) Biophys. J. 74, 3211–3216
11. Birk, D. E., and Trebold, R. L. (1986) J. Cell Biol. 103, 231–240
12. Makareeva, E., Mertz, E. L., Kuznetsova, N. V., Sutter, M. B., DeRidder, A. M., Cabral, W. A., Barnes, A. M., McBride, D. J., Marini, J. C., and Leikin, S. (2008) J. Biol. Chem. 283, 4787–4798
13. Brodsky, B., and Ramshaw, J. A. M. (1997) Matrix Biol. 15, 545–554
14. Hulmes, D. J. S., Miller, A., White, S. W., Timmins, P. A., and Berteth-Colominas, C. (1980) Int. J. Biol. Macromol. 2, 338–346
15. Orgel, J. P., Irving, T. C., Miller, A., and Wess, T. J. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 9001–9005
16. Gelman, R. A., Poppke, D. C., and Piez, K. A. (1979) J. Biol. Chem. 254, 11741–11745
17. Snowdon, J. M., and Swann, D. A. (1979) Biochem. Biophys. Acta 580, 372–381
18. Helseth, D. L., Jr., Lechner, J. H., and Veis, A. (1979) Biopolymers 18, 3005–3014
19. Helseth, D. L., Jr., and Veis, A. (1981) J. Biol. Chem. 256, 7118–7128
20. Capaldi, M. J., and Chapman, J. A. (1982) Biopolymers 21, 2291–2313
21. Jones, E. Y., and Miller, A. (1987) Biopolymers 26, 463–480
22. Prockop, D. J., and Fertala, A. (1998) J. Biol. Chem. 273, 15598–15604
23. Orgel, J. P., Wess, T. J., and Miller, A. (2000) Structure 8, 137–142
24. Cabral, W. A., Fertala, A., Green, L. K., Körkkö, J., Forlino, A., and Marini, J. C. (2002) J. Biol. Chem. 277, 4215–4222
25. Malone, J. P., and Veis, A. (2004) Biochemistry 43, 15358–15366
26. Kuznetsova, N., and Leikin, S. (1999) J. Biol. Chem. 274, 36083–36088
27. McDowall, M. A. (1970) Eur. J. Biochem. 14, 214–221
28. Carne, A., and Moore, C. H. (1978) Biochem. J. 173, 73–83
29. Morimoto, K., Furtuta, E., Hashimoto, H., and Inouye, K. (2006) J. Biochem. 139, 1065–1071
30. Morimoto, K., Kunii, S., Hamano, K., and Tonomura, B. (2004) Biosci. Biotechnol. Biochem. 68, 861–867
31. Morimoto, K., Kawabata, K., Kunii, S., Hamano, K., Saito, T., and Tonomura, B. (2009) J. Biochem. 145, 677–684
32. Yano, M., Nagai, K., Morimoto, K., and Miyamoto, H. (2007) Biochem. Biophys. Res. Commun. 362, 158–163
33. Nagai, K., Yotsukura, N., Ikegami, H., Kimura, H., and Morimoto, K. (2008) Electrophoresis 29, 672–681
34. Nakayama, T., Shirane, J., Hieshima, K., Shibano, M., Watanabe, M., Jin, Z., Nakagubo, D., Saito, T., Shimomura, Y., and Yoshie, O. (2006) Virology 350, 484–492
35. Netzelt-Arnett, S., Fields, G., Birkedal-Hansen, H., Van Wart, H. E., and Fields, G. (1991) J. Biol. Chem. 266, 6747–6755
36. Williams, B. R., Gelman, R. A., Poppke, D. C., and Piez, K. A. (1979) J. Biol. Chem. 254, 180–186
37. Malone, J. P., George, A., and Veis, A. (2004) Proteins 54, 206–215