Trimerization of the Amino Propeptide of Type IIA Procollagen Using a 14-Amino Acid Sequence Derived from the Coiled-Coil Neck Domain of Surfactant Protein D*

Received for publication, March 7, 2002, and in revised form, July 17, 2002 Published, JBC Papers in Press, August 22, 2002, DOI 10.1074/jbc.M202257200

Audrey McAlinden‡, Erika C. Crouch§, James G. Bann¶, Pengnian Zhang§, and Linda J. Sandell¶

From the Departments of ‡Orthopaedic Surgery, §Pathology and Immunology, and ¶Biochemistry and Molecular Biophysics, Washington University School of Medicine at Barnes-Jewish Hospital, St. Louis, Missouri 63110

The folding of a collagen triple helix usually requires the presence of additional sequences that contribute to the association and correct alignment of the collagen chains. We recently reported that the C-terminal neck and lectin domains of a collagenous C-type lectin, rat pulmonary surfactant protein D (SP-D), are sufficient to drive the trimerization of a heterologous type IIA pro-collagen amino propeptide sequence. However, the conformation of the resulting trimeric IIA propeptide and the specific contributions of the SP-D sequence to trimerization were not elucidated. In the present study, we show that trimerization of the fusion protein is associated with correct folding of the collagen helix within the IIA propeptide domain (as assessed by circular dichroism) and that the constituent chains are hydroxylated. Chemical cross-linking and analytical ultracentrifugation showed that the IIA amino-propeptide retains its trimeric configuration even after proteolytic removal of the SP-D domains. By contrast, IIA amino-propeptides synthesized without fusion to the neck or lectin domains are assembled exclusively as monomers. To localize the trimerization sequence, mutant chimeric cDNA constructs were designed containing premature termination codons within the coiled-coil neck domain. A short, 14-amino acid sequence corresponding to the first two heptad repeats of the neck domain was sufficient to drive the trimeric association of the IIA amino-propeptide α-chains. However, deletion of the collagen domain resulted in the secretion of monomers. These studies demonstrate that two heptad repeats are sufficient for trimeric association of the propeptide but indicate that cooperative interactions between the coiled-coil and collagen domains are required for the formation of a stable helix.

The type IIA NH2-propeptide is encoded by eight exons; the translated protein consists of a short globular domain, a 69-amino acid von Willebrand factor type C cysteine-rich domain, a minor collagen triple-helical domain containing 26 GXX repeats, and a short telopeptide domain, which links the minor collagen domain to the major collagen triple helix (1, 2). Trimerization of most fibrillar collagens is dependent on the globular COOH-propeptide for the recognition and association of the three polypeptide chains resulting in registered nucleation of triple helix formation in a zipper-like fashion from the C to N terminus (3–7). Functions proposed for procollagen NH2-propeptides include the regulation of collagen fibrillogenesis (8) and a feedback control of net collagen biosynthesis (9–11). Recently, it was proposed by our laboratory (12) and by others (13) that the NH2-propeptide of type IIA procollagen regulates growth factor activity in the extracellular matrix. Trimeric assembly of fibrillar NH2-propeptides affects protein valency and stability, which appear important for function in vivo. This emphasizes the importance of a procollagen COOH-propeptide, or indeed other protein domains with similar function, to drive this trimerization process.

Pulmonary SP-D1 is predominantly assembled as dodecamers, consisting of four trimeric subunits cross-linked by disulfide bonds (14, 15). Each SP-D subunit contains an N-terminal coiled-coil domain, an uninterrupted triple-helical collagen domain consisting of 59 GXX repeats, a trimeric coiled-coil neck domain, and a C-type lectin carbohydrate recognition domain (CRD). Trimerization of SP-D subunits and subsequent oligomerization of these trimeric subunits to form higher order multimers results in increased valency of the CRD, an essential prerequisite for high affinity ligand binding (16). The neck domain of SP-D is the unit responsible for driving the trimerization of the three polypeptide chains of SP-D. It was demonstrated that a 35-amino acid sequence containing the human neck sequence was sufficient to form stable, non-covalent, trimeric complexes in vitro (17). The same sequence was found to be important for the association of the three CRDs of human SP-D; CRDs synthesized in prokaryotic cells without this neck domain were assembled as monomers (18).

The sequence of coiled-coil domains is characterized by a 7-residue (heptad) repeat commonly denoted (abcdefg)n in which positions a and d are primarily occupied by hydrophobic residues, positions e and g are primarily occupied by charged residues, and positions b, c and f are primarily occupied by polar or charged residues (19). The crystal structure of the neck and lectin domain of human SP-D has been solved, and the coiled-coil region was visualized as a stretch of greater than 28 amino acids (Arg208–Pro235) consisting of 8 helical turns (20).

Based on the earlier work of Harbury et al. (21), Reid and

---

* This work was supported by National Institutes of Health Grants AR-36994 (to L. S.), HL-29594 (to E. C.), and HL-44015 (to E. C.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

To whom correspondence should be addressed: Dept. of Orthopaedic Surgery, Washington University School of Medicine at Barnes-Jewish Hospital, 660 South Euclid Ave., St. Louis, MO 63110. Tel.: 314-454-7800; Fax: 314-454-5900; E-mail: sandell@msnotes.wustl.edu.

**The abbreviations used are: SP-D, surfactant protein D; BS2, bis(sulfo succinimidyl)suberate; CRD, carbohydrate recognition domain; MMP, matrix metalloproteinase; PBS, phosphate-buffered saline; CHO, Chinese hamster ovary; EK, enterokinase.**
co-workers (17) suggested that the presence of valine at the d positions favors the trimeric assembly of human SP-D. It was further suggested that the unusual fourth heptad, which contains Phe<sup>225</sup> and Tyr<sup>228</sup> in the a and d positions, respectively, might serve to initiate trimerization (20). However, no valine residues are found in the neck of rat SP-D. In addition, we observed that deletion of the conserved fourth heptad repeat does not prevent trimerization of recombinant rat SP-D secreted by mammalian cells (22). On the other hand, internal deletions of residues 207–214 or 214–221 within the neck domain were found to block trimerization and indicated that sequences N-terminal to Phe<sup>225</sup> were required for trimerization. These studies also demonstrated that trimeric assembly of the NH<sub>2</sub>-propeptide of type IIA procollagen could be initiated by the C-terminal neck and lectin domains of SP-D. In doing so, the chain association function of the type IIA procollagen COOH-propeptide was replaced by these SP-D domains. However, the conformation of the resulting trimeric IIA complex and the specific contributions of the C-terminal SP-D sequence to trimerization were not elucidated.

The present study describes the molecular structure of the resulting IIA NH<sub>2</sub>-propeptide and establishes, for the first time, that a short, amphipathic helical sequence derived from the rat SP-D neck domain can also drive the trimerization of a fibrillar collagen NH<sub>2</sub>-propeptide. This sequence consists of the first two heptad repeats of the neck domain, which is in agreement with our previous deletional mutagenesis studies showing that N-terminal regions of the neck domain are important for initiating trimerization (22). There have been only a few reports on the use of oligomerization domains to trimerize heterologous collagenous sequences (23). To our knowledge, this is by far the shortest sequence found to permit trimerization of a collagenous molecule and the first to demonstrate the use of a heterologous trimerization cassette to support the normal folding of a procollagen propeptide domain. High levels of a correctly folded IIA NH<sub>2</sub>-propeptide were produced using this system, which will enable us to study its biological function in vitro. Establishing a minimum sequence of the SP-D neck domain that can drive trimerization without inclusion of the functional SP-D lectin domain will allow us to study the function of the trimeric IIA propeptide in vivo. Knowledge gained from these findings may be applied to produce other procollagen propeptides or indeed other collagenous proteins for functional studies.

**EXPERIMENTAL PROCEDURES**

**Expression of IIA/SP-D Fusion Protein in Chinese Hamster Ovary (CHO)-K1 cells**—A chimeric construct was synthesized by linking the cDNA encoding the NH<sub>2</sub>-propeptide of type IIA procollagen (see Fig. 1) to the cDNA encoding the neck + CRD of rat SP-D (see Fig. 2). This chimeric construct and the resulting fusion protein was named IIA/SP-D. The cDNA encoding exons 1–8 of human type IIA procollagen NH<sub>2</sub>-propeptide were amplified by reverse transcription-PCR from RNA that had been isolated from articular chondrocytes in culture. Specific upstream and downstream primers were designed from the pro-α1 type II collagen complete coding sequence (GenBank™ accession number: L10347). The IIA/SP-D chimeric construct was made by overlap extension PCR. Briefly, the complete coding sequence of the IIA NH<sub>2</sub>-propeptide (using oligomer A, cagcactgtcagctggctcg) with the 5′-end, shown in bold) and a 3′ sequence homologous to a region of the neck domain of rat SP-D (using oligomer B, cagcactgtcagctggctggctgt) was amplified by PCR for 25 cycles at an annealing temperature of 52 °C. The same conditions were used to amplify the neck + CRD of rat SP-D containing a 5′ sequence homologous to a region of the IIA CRD (using oligomer C, aggacactgtcagctggctcg) and a 3′-ECRI site (using a T7-specific downstream oligonucleotide). cDNA products from the two PCR amplifications were combined, and overlap extension PCR was carried out for 30 cycles at an annealing temperature of 55 °C using oligomers A and T7. The resulting chimeric construct was digested with EcoRI and subcloned into pGEM-3Z, and the orientation of the subcloned insert was confirmed by restriction mapping and DNA sequencing.

IIA/SP-D cDNA was excised from pGEM-3Z by EcoRI digestion and ligated into the multiple cloning site of pE14 (24) distal to the cytomegalovirus promoter/enhancer and proximal to the glutamine synthetase gene. CHO cells (CHO-K1; ATCC CCL-61) were transfected with pE14-IA/SP-D using LipofectAMINE (In Vitrogen) and grown in selection Glasgow’s minimum essential medium (In Vitrogen) containing 10% dialyzed fetal bovine serum and the glutamine synthetase inhibitor, methionine sulfoxamine (25–50 μM) for 2–3 weeks. Stable clones were obtained as described by Crouch et al. (25) for the expression of recombinant rat-SP-D. To assess the importance of the trimerizing neck domain, control pE14 constructs were synthesized consisting of cDNA encoding the full-length IIA NH<sub>2</sub>-propeptide, devoid of cDNA encoding the neck and lectin domains. This construct was used in transient transfections of CHO cells using LipofectAMINE reagent.

Detection and Purification of IIA/SP-D Fusion Protein—Media from transiently transfected CHO cells were screened for the presence of the fusion protein by an enzyme-linked immunosassay using rabbit anti-human exon 2 (IIA) antibody (26), chicken IgY anti-human Exon3-8 antibody (27), or rabbit anti-rat SP-D antibody (27). Immunopositive proteins labeled with rabbit-horseradish peroxidase secondary antibodies were detected by enhanced chemiluminescence using SuperSignal™ chemiluminescent substrate (Pierce). Clones expressing the IIA/SP-D fusion protein were selected and cultured further by exposure to 50 μM methionine sulfoxamine, and the resulting conditioned medium was dialysed against Tris-buffered saline, pH 7.5, containing 10 mM EDTA. CaCl<sub>2</sub> (20 mM) was added to the dialyzed medium, and IIA/SP-D was subsequently purified by maltosyl-agarose chromatography (25). Because the interaction of the CRD with calcium is calcium-dependent (27), IIA/SP-D was eluted from the column with Tris-buffered saline, 10 mM EDTA, pH 7.5. Eluted fractions were analyzed by SDS-polyacrylamide gel electrophoresis, silver staining, and Western blotting.

**Collagenase Digestion of IIA/SP-D—Bacterial collagenase was purified by gel filtration chromatography using crude collagenase as the starting material (Worthington Biochemical Corp. (28). IIA/SP-D or rat SP-D (5 μg) in Tris-buffered saline, pH 7.5, was digested with purified bacterial collagenase (1 μg) containing CaCl<sub>2</sub> (20 mM) and N-ethylmalaeimide (5 mM) overnight at 37 °C. Fresh collagenase (1 μg) was added for a further 3 h at 37 °C followed by EDTA (4 mM) to stop the reaction. An aliquot (5 μg) of digested and undigested IIA/SP-D or rat SP-D was electrophoresed through a 4–20% SDS-polyacrylamide gel to confirm collagenase digestion. The major collagenase-resistant products were detected by Coomassie Blue staining and subjected to N-terminal amino acid sequencing. Collagenase-digested IIA/SP-D or SP-D was transferred to Sequi-Blot polyvinylidene difluoride membrane (Bio-Rad), stained with Coomassie Blue, excised, and sequenced on an ABI 473A protein sequencer equipped with model 610A data analysis software.

**Purification of IIA NH<sub>2</sub>-Propeptide, MMP-9, or Enterokinase Cleavage of Wild-type or Mutant IIA/SP-D Fusion Protein—**Approximately 100 μg of wild-type IIA/SP-D fusion protein was digested overnight at 37 °C with human recombinant MMP-9 (gift from Dr. R. Senior, Pulmonary and Critical Care Division, Washington University in St. Louis, MO) at an enzyme-substrate ratio of 1:100. MMP-9 cleaves within the telopeptide domain of the IIA propeptide on either side of Glu<sup>157</sup> and Met<sup>177</sup> (27). Since MMP-9 has two cleavage sites within the telopeptide and cleavage is not always 100% efficient, we proceeded to synthesize a mutant IIA/SP-D chimeric construct containing an enterokinase cleavage site in the exon 8-encoded telopeptide. Using the QuikChange™ site-directed mutagenesis kit (Stratagene), two ODN primers were designed to change the DNA sequence encoding amino acids 161–165 in exon 8 of the wild-type IIA NH<sub>2</sub>-propeptide (161GFDERK165) to one that encodes the KE cleavage site (161DDDKRK165). Stable CHO cell lines producing this mutant fusion protein (IIA/EK/SP-D) were produced as described above. Approximately 0.001% w/v of the CHO cell line (New England Cloned Cell Line Service) was added to purified IIA/SP-D protein overnight at room temperature.

Cleavage by MMP-9 or enterokinase was confirmed by gel electrophoresis, silver staining, and immunoblotting using antibodies specific for the IIA (exon 2) domain or the CRD of SP-D. CaCl<sub>2</sub> was added to the cleaved products before applying to a maltosyl-agarose column to separate the IIA NH<sub>2</sub>-propeptide (from flow-through) from the IIA CRD of SP-D (present in the EDTA eluate).

**Chemical Cross-linking—**Covalent cross-linking was performed using bis-(sulfosuccinimidyl) suberate (BS3; Pierce) at an enzyme:substrate ratio of 1:100. MMP-9 cleaves the IIA NH<sub>2</sub>-propeptide and the specific contributions of the C-terminal SP-D sequence to trimerization were not elucidated.

The present study describes the molecular structure of the resulting IIA NH<sub>2</sub>-propeptide and establishes, for the first time, that a short, amphipathic helical sequence derived from the rat SP-D neck domain can also drive the trimerization of a fibrillar collagen NH<sub>2</sub>-propeptide. This sequence consists of the first two heptad repeats of the neck domain, which is in agreement with our previous deletional mutagenesis studies showing that N-terminal regions of the neck domain are important for initiating trimerization (22). There have been only a few reports on the use of oligomerization domains to trimerize heterologous collagenous sequences (23). To our knowledge, this is by far the shortest sequence found to permit trimerization of a collagenous molecule and the first to demonstrate the use of a heterologous trimerization cassette to support the normal folding of a procollagen propeptide domain. High levels of a correctly folded IIA NH<sub>2</sub>-propeptide were produced using this system, which will enable us to study its biological function in vitro. Establishing a minimum sequence of the SP-D neck domain that can drive trimerization without inclusion of the functional SP-D lectin domain will allow us to study the function of the trimeric IIA propeptide in vivo. Knowledge gained from these findings may be applied to produce other procollagen propeptides or indeed other collagenous proteins for functional studies.
sodium citrate, pH 5.5, were added to each recombinant protein for 1 h at room temperature. Addition of SDS-PAGE loading buffer containing Tris-HCl (0.5 x) inhibited the reaction. Samples were boiled for 5 min prior to SDS-PAGE, which was carried out in the absence of sulfhydryl reduction. Cross-linked proteins were identified by silver staining or immunolocalization using anti-IIA (exon 2) polyclonal antisera.

**Circular Dichroism and Determination of IIA NH2-Propeptide Melting Temperature**—Approximately 50 µg of the IIA NH2-propeptide (0.2 mg/ml in PBS, pH 7.5), purified by cleavage of the mutant IIA/EEK/SP-D fusion protein containing the enterokinase cleavage site, was analyzed by CD spectroscopy. A Jasco J715 spectropolarimeter with a thermostatted quartz cell and a path length of 1 cm was used, and the spectrum was recorded at 5 °C between 180 and 260 nm. To determine the melting temperature of the IIA NH2-propeptide, the spectrum was monitored at 225 nm from 5 to 70 °C.

**Analytical Ultracentrifugation**—Equilibrium sedimentation experiments were performed using a Beckman Optima XL-A analytical ultracentrifuge using a six-channel centerpiece in an AN-60 Ti rotor. The IIA NH2-propeptide, purified from enterokinase cleavage of the IIA/EEK/SP-D mutant protein in PBS (pH 7.5), was analyzed at three concentrations: 0.2, 0.4, and 0.8 mg/ml. Experiments were performed at two speeds (20,000 and 28,000 rpm) at a temperature of 20 °C and wavelength of 280 nm. Data were fitted using WinNonLin V1.035 (www.ucc.uconn.edu/~wwwbiote/UAF.html), and a partial specific volume of 0.73 cm3/g was used for determining the molecular weight.

**RESULTS**

**Purification of IIA/SP-D Fusion Protein**—A chimeric gene construct was synthesized consisting of cDNA encoding the full-length type IIA NH2-propeptide (exons 1–8; Fig. 1) fused to the cDNA encoding the neck domain and lectin domain of SP-D. The cDNA of SP-D, the chimeric construct, and the predicted structure of the resulting fusion protein, named IIA/SP-D, are shown in Fig. 2. IIA/SP-D was efficiently purified from all other contaminating proteins present in the conditioned medium of stably transfected CHO cells after maltosyl-agarose chromatography (Fig. 3A). The monomer protein showed an apparent molecular mass of 45 kDa in the absence of sulfhydryl reduction when compared with globular protein standards used in this gel system. Interestingly, a small population of stable trimers of IIA/SP-D, resistant to SDS treatment and boiling prior to gel electrophoresis, was also visualized. Similar stable trimers of the type I procollagen NH2-propeptide have been detected from bone (29).

Immunoblotting of the EDTA-eluted protein with anti-IIA, anti-IIIα3–8, or anti-SPD polyclonal antisera confirmed identification of IIA/SP-D. Results were identical with all three antibodies, and Fig. 3B shows the immunopositive IIA/SP-D bands after detection with the anti-Exon3–8 antibody. The fusion protein migrated more slowly after sulfhydryl reduction due to unfolding of the looped structure created by the formation of the two intrachain disulfide bonds present within the lectin domain of SP-D. Although the type I IIA NH2-propeptide domain is predicted to contain five intrachain disulfide bonds, the loops are comparatively small, and disruption of these bonds did not alter the electrophoretic migration of the protein (results not shown). However, disruption of the cysteine pairs within the IIA NH2-propeptide altered the structure of the exon 2-encoded domain such that recognition of the epitope by the anti-IIA antibody was affected (results not shown). All 10 cysteine residues in this domain are paired because reaction of IIA/SP-D with Ellman’s reagent (Pierce) showed no quantifiable yellow-colored product as would be expected in the presence of free sulfhydryl groups. This suggests the presence of a very intricately folded domain since the 10 cysteine residues within the type I IIA NH2-propeptide are arranged in close proximity to each other (Fig. 1).

**Analysis of the IIA NH2-Propeptide Collagen Domain**—To investigate the structure of the recombinant IIA NH2-propeptide, IIA/SP-D fusion protein was digested with purified bacterial collagenase, and the major collagenase-resistant bands were characterized by N-terminal sequencing. SP-D, which

---

**Fig. 1.** Amino acid sequence of human type IIA procollagen NH2-propeptide. The sequence begins at the signal peptide cleavage site, numbered as the first amino acid (Q). Arrows indicate exon (E) boundaries. The cysteine-rich, von Willebrand factor type C domain encoded by exon 2 is shown in italics. Underlined amino acids in the region encoded by exons 3–7 denote the minor collagen domain containing 26 GXY repeats and a 4-amino acid interruption between exons 4 and 5. The telopeptide domain connects the propeptide to the major procollagen triple-helical domain and is encoded by exon 8.

---

**FIG. 1.** Amino acid sequence of human type IIA procollagen NH2-propeptide. The sequence begins at the signal peptide cleavage site, numbered as the first amino acid (Q). Arrows indicate exon (E) boundaries. The cysteine-rich, von Willebrand factor type C domain encoded by exon 2 is shown in italics. Underlined amino acids in the region encoded by exons 3–7 denote the minor collagen domain containing 26 GXY repeats and a 4-amino acid interruption between exons 4 and 5. The telopeptide domain connects the propeptide to the major procollagen triple-helical domain and is encoded by exon 8.

---

**FIG. 2.** IIA/SP-D was efficiently purified from all other contaminating proteins present in the conditioned medium of stably transfected CHO cells after maltosyl-agarose chromatography (Fig. 3A). The monomer protein showed an apparent molecular mass of 45 kDa in the absence of sulfhydryl reduction when compared with globular protein standards used in this gel system. Interestingly, a small population of stable trimers of IIA/SP-D, resistant to SDS treatment and boiling prior to gel electrophoresis, was also visualized. Similar stable trimers of the type I procollagen NH2-propeptide have been detected from bone (29).
contains its own collagen domain, was included as a control. As shown in Fig. 4A (protein bands 1–3), most of the GXY collagen domain in IIA/SP-D and SP-D was digested (Fig. 4B). In addition, amino acid analysis of IIA/SP-D showed that there were 8 hydroxyproline residues in the collagen domain of the IIA NH2-propeptide. There are 11 potential sites for proline hydroxylation (GXP), but it is not known what percentage of prolines is hydroxylated in the native type II propeptide. To further determine the trimeric configuration of the IIA NH2-propeptide, we chose to purify the propeptide from the neck/CRD of SP-D. This was done by cleavage of the wild-type IIA/SP-D protein with MMP-9 or by digestion of the mutant fusion protein (IIA/EK/SP-D) synthesized with an enterokinase cleavage site within the exon 8-encoded telopeptide domain (Fig. 5A). After cleavage, the digested protein fragments were applied to a maltosyl-agarose column, which binds to the trimeric neck/CRD fragments. The IIA NH2-propeptide was present in the flow-through, and the SP-D fragments were then eluted with EDTA (Fig. 5B).

To confirm that the IIA NH2-propeptide contained a correctly folded collagen triple helix, the propeptide purified by enterokinase cleavage of IIA/EK/SP-D was analyzed by CD spectroscopy. The CD spectrum of a collagen triple helix is characterized by a small positive peak at 220–225 nm, a cross-over at 213 nm, and a trough at 197 nm (30). Fig. 6A shows a large positive ellipticity at 225 nm, indicative of a collagen triple helix. The IIA propeptide was heated to 70 °C, and the CD spectrum was monitored at 225 nm. Fig. 6B shows that the mean residue ellipticity (θ) decreased with increasing temperature and that the melting temperature of the collagen triple helix was ~42 °C. Final confirmation that the IIA propeptide exists as a trimer in solution was achieved by analytical ultracentrifugation, using the sedimentation equilibrium approach, to calculate the molecular weight. The expected molecular weight of the trimeric propeptide was estimated using a ProtParam program (us.expasy.org/tools/protparam.html) and was found to be 50,118 g/mol. The actual molecular weight calculated using the sedimentation equilibrium method was 50,838 g/mol.

**A Trimerization Domain Is Necessary for the Production of a Correctly Folded IIA NH2-Propeptide**—Chemical cross-linking was used to examine the state of oligomerization of the IIA collagen domain. In particular, cross-linking profiles were compared for: 1) the wild-type IIA/SP-D fusion protein, 2) the IIA NH2-propeptide purified after MMP-9 cleavage of the fusion protein, and 3) the IIA NH2-propeptide synthesized without fusion to the neck/CRD domains of SP-D. As shown in Fig. 7,
cross-linking resulted in the dose-dependent appearance of IIA/SP-D trimers (T) through a dimeric (D) intermediate. As expected, the isolated IIA NH2-propeptide showed a similar cross-linking pattern. By contrast, the IIA NH2-propeptide expressed in the absence of SP-D sequence showed no evidence of cross-linked dimers or trimers, indicating the secretion of monomers (M).

A 14-amino Acid Sequence of the Coiled-Coil Neck Domain Can Drive the Trimerization of the IIA NH2-Propeptide—The trimerization domain of rat SP-D is a coiled-coil structure that consists of four heptad repeats as depicted in Fig. 8. To further assess the relative contributions of subregions of the neck domain, IIA/SP-D truncation mutants were synthesized by introducing premature stop codons within the coiled-coil neck domain to produce the IIA NH2-propeptide attached to one, two, or three contiguous heptad repeats. Two additional mutant IIA/SP-D proteins were generated as controls. One contained a stop codon at the first amino acid of the neck domain (Asp203) or at the final residue (Gly 237) of the 35-amino acid sequence originally identified as the SP-D trimerization unit (17, 18) (Fig. 8). Each mutant protein was covalently cross-linked, and the presence of protein monomers, dimers, or trimers was detected by immunoblotting using the anti-IIA antibody. Fig. 9 shows that the IIA NH2-propeptide lacking the neck domain sequence (mIIA-203) or fused to the first heptad repeat (mIIA-211) were secreted as monomers. However, truncated fusion proteins containing two or three heptad repeats (mIIA-218 and mIIA-225, respectively) showed trimeric assembly. The IIA NH2-propeptide attached to the 35-amino acid stretch of the coiled-coil neck (mIIA-237) was also secreted as a non-covalent trimer. However, lower concentrations of cross-linker (0.1–0.2 mM) were sufficient for detection of mIIA-237 trimers as compared with concentrations used to detect trimers of the other truncated mutant proteins (0.5–1 mM), and no dimeric intermediate was identified.

Cooperativity Exists between the IIA NH2-Propeptide Collagen Domain and the 14-amino Acid Sequence of the SP-D Coiled-Coil Neck Domain—Based on published literature (31), it was shown that a two heptad repeat coiled-coil sequence cannot form an autonomous folding unit. Thus, it is highly likely that cooperative interactions exist between the collagen domain and the short, 14-amino acid sequence of the SP-D trimerization domain to stabilize the truncated fusion protein (Figs. 8 and 9, mIIA-218). A collagen deletion construct was synthesized to produce a mutant protein consisting of exons 1, 2, and 8 of the IIA NH2-propeptide fused to two heptad repeats of the coiled-coil domain (mIIA-coll-218; Fig. 10A). Protein from conditioned medium of transiently transfected CHO cells was
Trimerization of Type IIA Procollagen Amino Propeptide

DISCUSSION

These studies demonstrate that a short, 14-amino acid sequence derived from the first two heptad repeats of the α-helical coiled-coil domain of rat SP-D can drive the trimerization of a heterologous procollagen NH2-propeptide sequence. Although IIA propeptides alone are secreted as monomers, a IIA/SP-D chimera with a truncated SP-D neck domain terminating at residue 211 or 203, containing one or no heptad repeats, respectively, were secreted as monomers. To our knowledge, this is the shortest sequence shown to support the trimerization of a collagen sequence.

In addition, we have shown that trimerization is accompanied by folding of the collagen triple helical domain and that, following cleavage from the SP-D sequence, the IIA NH2-propeptide retains its trimeric conformation. Amino acid analysis revealed that ∼80% of the potential proline residues in the Y position of the collagen sequence are hydroxylated, consistent with the formation of a stable triple helix. These levels of hydroxylation are comparable with those reported for the α1 chain of the NH2-propeptide of type I procollagen extracted from developing bone (29).

In addition, the melting temperature of the collagen helix within the recombinant propeptide from developing bone (29). In addition, the melting temperature of the collagen helix within the recombinant propeptide from developing bone (29) was similar to other comparably hydroxylated collagens.

The requirements for collagen trimerization and folding vary with the collagen type. Generally, fibrillar collagens and type IV collagen require the presence of globular sequences C-terminal to the triple-helical domain to initiate chain registration (3–5, 32). However, trimerization of type XII collagen is dependent on specific post-translational modifications of the collagen domain (33), whereas chain association of the membrane-associated collagen, type XIII, occurs in the N-terminal region (34). Refolding experiments on collagen type III indicated that interchain disulfide bridges at the C terminus of the triple helix were sufficient to function as a nucleus for the refolding of the triple helix (6). These findings suggest that the sequences required for driving collagen trimerization can be manipulated as also exemplified by our ability to trimerize a procollagen amino propeptide using the α-helical coiled-coil domain of rat SP-D.

There is still little precedent for the use of heterologous sequences to drive the trimerization of collagen sequences. Recently, Frank et al. (23) utilized the bacteriophage T4 fibrin foldon domain to synthesize a chimeric protein consisting of a synthetic collagen peptide (Pro-Pro-Gly)10 fused to the N-terminus of the foldon. The foldon domain (35), which consists of 27 amino acids and forms a β-propeller-like structure with a hydrophobic interior, was sufficient to drive the trimerization

Fig. 7. Covalent cross-linking of IIA/SP-D or IIA NH2-propeptides synthesized with or without the trimerization domain of SP-D. The transition from monomers (M) to trimers (T) through a dimer (D) intermediate with increasing concentrations of BS3 cross-linker is shown for IIA/SP-D. The same pattern is shown for the purified IIA NH2-propeptide that was synthesized attached to the neck and lectin domain of SP-D and then subsequently purified by MMP-9 treatment. * MMP-9-derived product not immunoreactive with either IIA or SP-D antisera. The IIA Western blot shows that the type IIA NH2-propeptide produced in transiently transfected CHO cells without the trimerization cassette of SP-D exists only as monomers in solution.

Fig. 8. Amino acid sequence of SP-D α-helical coiled-coil neck domains from different species and schematics showing mutant IIA/SP-D fusion proteins containing a premature stop codon within the coiled-coil domain. The amino acid sequence of the coiled-coil neck domain shows the presence of four contiguous heptad repeats. Positions a and d, generally represented by hydrophobic residues, are indicated. The schematic below shows the complete sequence of rat SP-D neck domain attached to the IIA NH2-propeptide at its N-terminal side. The coiled-coil sequence ends at the last proline residue (Pro218) and proceeds to the sequence encoding the CRD. Underlined amino acids represent locations where the codon was replaced by a premature stop site in the cDNA sequence. Each mutant (m) protein consists of the full-length IIA NH2-propeptide sequence fused to either one (mIIA-211), two (mIIA-218), or three heptad repeats (mIIA-225) of the coiled-coil neck domain. The IIA NH2-propeptide devoid of the neck sequence (mIIA-203) or attached to the full-length sequence reported previously to drive trimerization (mIIA-237) were included as controls. Amino acids labeled with asterisks (•) indicate residues that may participate in electrostatic interactions to stabilize the coiled-coil at its N-terminal end: Arg208 to Gln212 (i to i + 4 intrachain) and/or Asp203 to Arg208 (i to i + 5; g-e′ interchain).
and correct folding of the synthetic collagen domain. In addition, the thermal stability of the folded collagen peptide was increased by fusion to the foldon domain. Another recent study showed that the COOH-propeptide of type III procollagen could be replaced with a transmembrane domain without affecting the folding of the collagen triple helix (36).

The ability of a 14-amino acid sequence to direct trimerization is surprising. Previous studies have shown that a classical two heptad repeat coiled-coil sequence is unable to form an autonomous folding unit (31). Even the complete neck domain of SP-D is short as compared with many coiled-coil domains, which average 7 repeats or 14 helical turns for three-stranded coiled-coils (37). As indicated in the Introduction, the potential importance of \( \beta \)-branched side chains for determining the assembly of coiled-coils was emphasized by Harbury et al. (21). In that study, the occurrence of \( \beta \)-branched residues at the \( d \) position disfavored dimers, whereas these residues at the \( a \) position disfavored tetramers, and the presence of branched residues at both positions favored trimers. Given the occurrence of valine residues in the first three \( \alpha \) positions of the human SP-D neck sequence (Fig. 8), Hoppe et al. (17) suggested that this feature contributes to trimeric assembly. However, no \( \beta \)-branched amino acids occur in these positions in the rat sequence, and none of the other known SP-Ds or related collectins show a similar conservation of \( \beta \)-branched residues in this position. Studies using model peptides and surveys of known coiled-coils have identified residues that favor various oligomeric states (38–40). Interestingly, residues found in the \( a \) and \( d \) positions of SP-D are usually non-discriminatory with respect to oligomerization or favor dimer formation. For example, leucine, which is present in the \( d \) position of the first three heptad repeats of SP-D, marginally favors dimers over trimers. Consistent with these observations, analysis of both human and rat \( \alpha \)-helical coiled-coil sequences using MultiCoil (41) predicted a dimeric association. For example, dimer formation probability for the human SP-D coiled-coil sequence was ~90%, or 70% for the rat sequence, using the available windows of 21 residues.

Thus, it seems likely that other interactions contribute to the stability or oligomerization of the 14-amino acid sequence. In this regard, \( g-e' \) ion-dipole interactions can contribute to the stability and oligomerization of some \( \alpha \)-helical coiled-coils (42, 43). Although most discussions emphasize the effects of electrostatic interactions on stability, Beck et al. (44) recently showed that specific electrostatic interactions were required for trimerization of the considerably longer coiled-coil domain of the cartilage matrix protein. Inspection of the neck sequence of rat SP-D suggests the possible occurrence of an intrahelical ionic interaction (i to i + 4 spacing between Arg\(^{208} \) and Glu\(^{212} \)) and/or an interchain ionic interaction (i to i + 5 spacing between Asp\(^{206} \) and Arg\(^{208} \); g-e') (Fig. 8). Studies are underway to replace the Arg\(^{208} \) residue with Gln to investigate subsequent effects on trimeric assembly.

In any case, our finding that mIIA-218 is secreted as monomers, whereas IIA-218 is secreted as trimers, is consistent with contributions of the collagen domain to trimer stability and the findings of Su et al. (31). Thus, both the N-terminal heptad repeats of the neck of SPD and the IIA collagen sequence are required to form stable chimeric trimers. To our knowledge,

![Figure 9](image-url)

**Fig. 9.** Chemical cross-linking of IIA NH2-propeptides fused to different regions of the SP-D coiled-coil neck domain. To determine the minimum sequence of the coiled-coil domain that can function as a trimerization domain, increasing amounts of cross-linker (BS3) were added to each mutant protein. Western blotting and immunolocalization using the anti-IIA polyclonal antibody were used to detect the protein. IIA NH2-propeptides devoid of the coiled-coil neck domain (mIIA-203) or containing one heptad repeat of the neck domain (mIIA-211) were shown to exist only as monomers (M) in solution. However, for the IIA NH2-propeptides attached to either two (mIIA-218) or three (mIIA-225) heptad repeats, trimer (T) formation is noted through a dimer (D) intermediate with increasing amounts of BS3. The mutant protein consisting of the IIA propeptide attached to full-length neck sequence (mIIA-237) was more efficiently trimerized at lower concentrations of cross-linker than those used for the other truncated proteins and, in addition, no dimer intermediate was detected.

![Figure 10](image-url)

**Fig. 10.** Production and chemical cross-linking of a collagen deletion mutant protein. A, schematic showing the collagen deletion protein (mIIA-coll-218) consisting of exons 1, 2, and 8 of the IIA NH2-propeptide fused to the short, 14-amino acid sequence of the SP-D coiled-coil neck domain (represented by the diagonal-shaded box). B, IIA immunoblot showing the presence of the collagen deletion protein from conditioned medium of transiently transfected CHO cells. There was no detection of dimers or trimers after addition of the highest concentration of cross-linker (BS3, 2 mM). Without the collagen domain (encoded by exons 3–7), the truncated fusion protein exists as monomers in solution.
Trimerization of Type IIA Procollagen Amino Propeptide

this is the direct demonstration of a cooperative and mutually stabilizing interaction between a collagen domain and its non-collagenous trimerization domain.

We were intrigued by the finding that the mIIA-237 fusion protein reproducibly trimerizes but without a detectable dimeric intermediate. Trimerization was also more efficient, requiring less cross-linker than for the other truncation mutants. We speculate that this "all-or-none" cross-linking of mIIA-237 results from the presence of two contiguous sites for BS\(^8\) cross-linking at Lys\(^228\) and Lys\(^230\) within the fourth heptad repeat. Although this seems at odds with the observation that cross-linking of IIA/SP-D also proceeds through a dimeric intermediate, the three chains may not be within an equivalent environment as compared with the context of the intact neck + CRD domain. The crystal structure of the human SP-D neck + CRD shows a striking deviation from 3-fold symmetry involving the fourth heptad repeat with one of the three tyrosines at position 228 buried and the other two partially exposed (20). Thus, our findings are consistent with the possibility that this is the direct demonstration of a cooperative and mutually stabilizing interaction between a collagen domain and its non-collagenous trimerization domain.

ACKNOWLEDGMENTS—We thank Tom Broekelmann for N-terminal protein sequencing, Hans-Peter Bachinger for hydroxyproline analysis, Troy Schumacher for preliminary protein purifications, and Gloria Hoch for excellent assistance in preparation of this manuscript.

REFERENCES

1. Ryan, M. C., and Sandell, L. J. (1990) J. Biol. Chem. 265, 10334–10339
2. Su, M. W., Benson-Chanda, V., Visiting, H., and Ramírez, F. (1989) Genomics 4, 438–441
3. Lamande, S. R., Chessler, S. D., Golub, S. B., Byers, P. H., Chan, D., Cole, W. G., Silence, D. O., and Bateman, J. F. (1995) J. Biol. Chem. 270, 8642–8649
4. Doeg, K. J., and Fessler, J. H. (1986) J. Biol. Chem. 261, 8924–8935
5. Chessler, S. D., Wallis, G. A., and Byers, P. H. (1993) J. Biol. Chem. 268, 18218–18225
6. Bachinger, H. P., Bruckner, P., Timpl, R., Prockop, D. J., and Engel, J. (1980) Eur. J. Biochem. 106, 619–632
7. Bruckner, P., Eikenberry, E. F., and Prockop, D. J. (1981) Eur. J. Biochem. 118, 607–613
8. Fleischmajer, R. (1986) J. Invest. Dermatol. 87, 553–554
9. Wu, C. H., Dunyan, C. B., and Wu, G. Y. (1986) J. Biol. Chem. 261, 482–484
10. Paglia, L., Wilczek, J., de Leon, L. D., Martin, G. R., Horlein, D., and Muller, P. (1979) Biochemistry 18, 5030–5034
11. Westermeier, M., Kreig, T., Horlein, D., Glavieux, R. W., Fietzek, P., and Muller, P. K. (1979) J. Biol. Chem. 254, 7016–7023
12. Zhu, Y., Oganessian, A., Keene, D. R., and Sandell, L. J. (1999) J. Cell Biol. 144, 1069–1080
13. Larrain, J., Bachiller, D., Lu, B., Agius, E., Piccolo, S., and de Robertis, E. M. (2000) Development 127, 821–830
14. Crouch, E. C. (1996) Am. J. Respir. Cell Mol. Biol. 19, 177–201
15. Crouch, E. C. (1996) Biochim. Biophys. Acta 1302, 278–289
16. Crouch, E. (2001) Respir. Res. 1, 93–108
17. Hoppe, I. J., Barlow, P. N., and Reid, K. B. (1994) FEBS Lett. 344, 191–195
18. Kishore, U., Wang, J. Y., Hoppe, H. J., and Reid, K. B. (1996) Biochem. J. 318, 505–511
19. McLachlan, A. D., and Stewart, M. (1975) J. Mol. Biol. 98, 293–304
20. Hakkanson, K., Lim, N. K., Hoppe, H. J., and Reid, K. B. (1999) Structure 7, 255–264
21. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) Science 262, 1401–1407
22. Zhang, P., McInden, A. Li, S., Schumacher, T., Wang, H., Hu, S., Sandell, L., and Crouch, E. (2001) J. Biol. Chem. 276, 18692–18670
23. Frank, S., Kammerer, R. A., Meichling, D., Schulthess, T., Landwehr, R., Bann, J., Guo, Y., Lustig, A., Bachinger, H. P., and Engel, J. (2001) J. Mol. Biol. 308, 1081–1089
24. Ausubel, R. M., Brent, R., Kingston, R. E., Moore, S. S., Seidman, J. G., Smith, J. A., and Struhl, K. (1992) in Current Protocols in Molecular Biology (Ausubel, R. M., Brent, R., Kingston, R. E., Moore, S. S., Seidman, J. G., Smith, J. A., and Struhl, K., eds.) pp. 16.14–16.14, John Wiley & Sons, Inc., New York
25. Crouch, E., Chang, D., Rast, K., Persson, A., and Heuser, J. (1994) J. Biol. Chem. 269, 15080–15083
26. Oganessian, A., Zhu, Y., and Sandell, L. J. (1997) J. Histochem. Cytochem. 45, 1795–1802
27. Persson, A., Chang, D., and Crouch, E. (1990) J. Biol. Chem. 265, 5755–5760
28. Peterkofsky, B., and Diegelmann, R. (1971) Biochemistry 10, 988–994
29. Fisher, L. W., Robey, P. G., Tuross, N., Otsuka, A. S., Tepen, D. A., Esch, F. S., Shimasaki, S., and Termine, J. D. (1987) J. Biol. Chem. 262, 13457–13463
30. Goodman, M., Bhumralkar, Jefferson, A. E., Kwar, J., and Locardi, E. (1998) Biopolymers 47, 127–142
31. Su, J. Y., Hodges, R. S., and Kay, C. M. (1994) Biochemistry 33, 15501–15510
32. Delz, R., Engel, J., and Kuhn, J. (1988) Eur. J. Biochem. 178, 357–366
33. Marrasera, M., Snellman, A., Kivirikko, K. I., van der Rest, M., and Pihlajaniemi, T. (1996) J. Biol. Chem. 271, 29052–29058
34. Snellman, A., Tu, H., Vaisanen, T., Kvist, A. P., Huhtala, P., and Pihlajaniemi, T. (2000) EMBO J. 19, 5051–5059
35. Tao, Y., Strekelov, S. V., Mysyanzhinov, V. V., and Rossman, M. G. (1997) Structure 5, 789–798
36. Bulled, N. J., Dalley, J. A., and Lees, J. F. (1997) EMBO J. 16, 6694–6701
37. Brown, J. H., Cohen, C., and Parry, D. A. (1986) Proteins 26, 134–145
38. Triplet, B., Wagschal, K., Lavigne, P., Mant, C. T., and Hodges, R. S. (2000) J. Mol. Biol. 300, 377–402
39. Wagschal, K., Triplet, B., Lavigne, P., Mant, C. T., and Hodges, R. S. (1999) Protein Sci. 8, 2312–2329
40. Wollf, E., Kim, P. S., and Berger, R. (1997) Protein Sci. 6, 1179–1189
41. Hu, J. C., Newell, N. E., Tidor, B., and Sauer, R. T. (1993) Protein Sci. 2, 1072–1084
42. Kohn, W. D., Kay, C. M., and Hodges, R. S. (1998) J. Mol. Biol. 283, 993–1012
43. Beck, K., Gunem, J. E., Kamalawal, A., and Bachinger, H. F. (1997) EMBO J. 16, 3767–3777