Assembly of the Type 1 Procollagen Molecule: Selectivity of the Interactions between the α1(I)- and α2(I)-Carboxyl Propeptides†

Keith Alvares, Farida Siddiqui, James Malone, and Arthur Veis*

Division of Oral Biology, Northwestern University Dental School, 303 East Chicago Avenue, Chicago, Illinois 60611

Received September 10, 1998; Revised Manuscript Received February 3, 1999

ABSTRACT: Assembly of the heterotrimeric procollagen I molecule is initiated by interactions between the carboxyl propeptide domains of the completed nascent pro α chains. The [pro α1(I)]$_2$[pro α2(I)] heterotrimer is the predominant molecule, with much smaller amounts of stable [pro α1(I)]$_3$ homotrimer also being formed. However, the [pro α2(1)]$_3$ homotrimer has not been detected, raising questions as to the mechanism of chain assembly and why [pro α2(1)]$_3$ homotrimers are not formed. These questions have been examined here by expressing the intact and amino- or carboxyl-terminal truncated C-propeptides of the pro α chains recombinantly in bacteria and in a coupled transcription/translation reticulocyte lysate system. Their interactions were studied in vitro by binding analyses and in vivo by using the yeast two-hybrid system. The C-pro α1(1) interacted with itself, and with C-pro α2(I), as expected. Surprisingly, the C-pro α2(I) also interacted with itself, both in vitro and in vivo. While the interaction of C-pro α2(I) with itself and C-pro α1(I) in vitro was strong, these interactions were weaker in vivo. Deletion of 36 amino acids from the C-terminal domain of C-pro α1 had no effect on its binding to intact self or intact C-pro α2, but the same deletion in C-pro α2 completely abolished its binding to intact C-pro α2 and to C-pro α1. Comparable N-terminal deletions in C-pro α1 or C-pro α2 diminished, but did not abolish, their binding to intact C-pro α1 and C-pro α2. In the yeast two-hybrid system, C-pro α2 interacted with itself more weakly than with C-pro α1. Molecular modeling and circular dichroism analyses showed that C-pro α1 and C-pro α2 have different folded structures and stability. Studies with antibodies specific to the C-pro α1 and α2 peptides showed them to precipitate different, specific, and distinct cell proteins from fibroblast lysates. The C-pro α2(I) antibody complexed with more cell proteins. We hypothesize that the lack of pro α2(I) homotrimers is not due to the inability of the C-pro α2(I) to interact with itself, but rather to the competing presence of other cell proteins. The specificity of these interactions may reside in conformational differences in N- and C-terminal sequences of the two propeptides or in their different folding patterns.

Collagens are the major structural components of the extracellular matrix. Type 1 collagen molecules are normally composed of two α1-chains and one α2-chain (*1*). The polypeptide chains are synthesized individually with identical length triple helix (G-X-Y)$_{338}$ structural domains, but the α1- and α2-chains have distinctive amino (N-) and carboxyl (C-) terminal extension telo- and pro-peptides. Although the N-propeptide domains are synthesized first, molecular assembly begins with the in-register association of the C-propeptides. It is believed the C-propeptide complex is stabilized by the formation of interchain disulfide bonds, but the necessity of disulfide bond formation for proper chain registration is not clear (*1, 2*). The triple helix is then propagated from the associated C-termini to the N-termini in a zipper-like manner (*3*, *4*). The [pro α1(I)]$_2$[pro α2(I)] heterotrimer is the predominant molecule formed. Small amounts of stable [pro α1(I)]$_3$ homotrimer are also formed and are present at variable levels (*5*, *6*) in different tissues. However, neither a [pro α1(I)][pro α2(I)]$_2$ heterotrimer nor a stable [pro α2(I)]$_3$ homotrimer has been detected. Thus, chain selection for molecular assembly has considerable specificity. Efficient procollagen I heterotrimer assembly appears to require that appropriate sets of elongating chains be inserted into the same ER compartment, and that a mechanism for correct chain recognition must exist within the ER. An open question in this biosynthetic pathway remains as to how the different gene products are selected, aligned, and subsequently folded into the triple helix (*7*). Most of our understanding of how procollagen assembles within cells comes from studies of the cells grown in culture (*1*, *2*, *8*). The formation of the triple helix of collagen has also been extensively studied using the renaturation of denatured collagen (*9*). These approaches have yielded valuable insights into the assembly of collagen; however, they have not provided a definitive analysis of the in vivo mechanism of molecular assembly.

To understand the initial events in the recognition of the three chains of type I procollagen, we have recombinantly expressed the C-propeptides of both pro α chains and studied their interactions in vitro, and in vivo using the yeast two-hybrid system. We show that, surprisingly, C-pro α2(I) is capable of binding to itself with substantial affinity. However, the binding of C-pro α2(I) with itself is lower in vivo than

† This work has been supported by the National Institute for Arthritis and Musculoskeletal Diseases (Grant AR-13921).

* Corresponding author: Northwestern University, 303 E. Chicago Ave., Chicago, IL 60611. Phone: 312-503-1355. FAX: 312-503-2544. E-mail: aveis@nwu.edu.

10.1021/bi9821824 CCC: $18.00
Published on Web 04/09/1999

in vitro, and this has important implications as to the possible role of other proteins in regulating the procollagen I self-assembly process. Deletion constructs of the propeptides show that a distinct domain in the C-terminal region of the C-pro α2(I) propeptide is necessary for interaction with C-pro α1(I) and with itself.

MATERIALS AND METHODS

Oligonucleotide Sequences. The sequences of the oligonucleotides used for PCR amplification of the various plasmid constructs are given below. Sequences underlined denote cloning sites added; the acronyms of the restriction enzymes are indicated on the right.

OL7	GCAGCCATATGGATGCCAATGTGGTTCGT	Nde1
OL8	CAGCCGGATCCTTACAGGAAGCAGACAGG	BamH1
OL9	GCAGCCATATGTATGACTTTGGTTACGAT	Nde1
OL10	GTGCCGGATCCTTATTTGAAACAGACTGG	BamH1
OL11	ATATGGCCATGGATGCCAATGTGGTTCG	Nco1
OL12	ATATGGCCATGGACTTTGGTTACGATGGA	Nco1
OL15	GCCGCTTCTAGATGGCCCGCACCTGCCGTG	Xba1
OL16	CCGGCTCTAGATGGATGCCAATGTGGTTC	Xba1
OL17	CCGGCGGATCCTTATTCAATCACTGTCTT	BamH1
OL18	GCCGCTCTAGATGGCTCGCACATGCCGTG	Xba1
OL19	GCCGGTCTAGATGGACTTTGGTTACGATG	Xba1
OL20	GGCGCGGATCCTTATTCAATGATTGTCTT	BamH1

Plasmid Construction. Sequence information from human pro α1(I) and pro α2(I) plasmids, from pHHα1 (7) and pHHα2 (7), respectively, was used to design primers to amplify the C-pro α1(I) and C-pro α2(I) domains or specific deletion constructs. The amplified products were then cloned using standard protocols (*10*). A summary of the different constructs is given below.

(A) Histidine-Tagged C-Pro α1(I) and C-Pro α2(I). Recombinant plasmids allowing expression of C-pro α1(I) and C-pro α2(I) as fusion proteins with six histidines at the amino terminal were constructed as follows. The C-pro α1 was amplified from pHHα1 (7) by the polymerase chain reaction using the forward primer OL7 and the reverse primer OL8. The C-pro α2 was amplified from pHHα2 (7) using the forward primer OL9 and the reverse primer OL10. The underlined sequences denote the *Nde*I and *Bam*HI restriction sites, respectively. The amplified products were then digested with *Nde*I and *Bam*HI and cloned in the *Nde*I, *Bam*HI site of the expression vector pET 15(b) to generate the plasmids pET-α1 and pET-α2. Recombinant plasmids were then cloned into the bacterial strain BL21(DE3). The C-pro α1(I) construct terminated at the junction of the propeptide and telopeptide. Because of cloning difficulties, the C-pro α2(I) was longer, and included an additional 12 amino acids of the C-telopeptide.

(B) 5′ Deletion of C-Pro α1(I) and C-Pro α2(I). PCR was employed to produce fragments lacking the first (N-terminal) 38 amino acids of C-pro α1(I) and 50 amino acids of C-pro α2(I), using primers OL15 and OL8 to amplify from pHHα1 and OL18 and OL10 to amplify from pHHα2. The amplified fragments were then cloned in pBluescript KS into the *Xba*I and *Bam*HI sites, generating the plasmids Δ5′C-proα1 and Δ5′C-proα2, respectively. The 5′ deletion constructs were of virtually identical length and left all the cysteine residues in C-pro α1(I) and C-pro α2(I) intact.

(C) 3′ Deletion of C-Pro α1(I) and C-Pro α2(I). Oligonucleotide pairs OL16, OL17 and OL19, OL20 were used as primers in a PCR to amplify fragments that lacked the last 36 amino acids of C-pro α1(I) and C-pro α2(I) from pHHα1 and pHHα2, respectively. The amplified fragments were then cloned in pBluescript KS into the *Xba*I and *Bam*HI sites to give the constructs Δ3′C-proα1 and Δ3′C-proα2. The deletion of 36 amino acids resulted in the loss of the last cysteine residue in both C-pro α1(I) and C-pro α2(I).

(D) Full-Length C-Pro α1(I) and C-Pro α2(I). The full-length C-pro α1(I) and C-pro α2(I) were amplified from pHHα1 and pHHα2 using the primer pairs OL16, OL8 and OL19, OL10, respectively. The resulting fragments were cloned in pBluscript KS.

Expression and Purification of C-Pro α1 and C-Pro α2. A single colony was selected and inoculated into 50 mL of Luria−Bertani broth (LB) and grown overnight at 30 °C. The next day, the culture was transferred to 500 mL of LB and grown at 30 °C until the OD_{600} was 0.6 (about 4 h). A portion of the uninduced cells was removed from the culture, and isopropyl thiogalactoside (IPTG) was then added to the remainder to a final concentration of 1 mM and incubation was carried on for 4 h. Samples were removed at various time points of induction. Cell lysates from the uninduced culture and at different time points of the induced culture were prepared and analyzed by SDS−PAGE (*11*) followed by Coomassie Blue staining. At the end of 4 h, the induced cells were fractionated into soluble and insoluble fractions and analyzed by SDS−PAGE. The bulk of the recombinant protein (>90%) accumulated in the insoluble fraction.

For the purification of the fusion proteins at the end of the 4 h induction, the cells were recovered by centrifugation at 5000*g* for 30 min. The cells were washed once with ice-cold 50 mM Tris-HCl, pH 8.0, and resuspended in 50 mL of 50 mM Tris-HCl, pH 8.0. Lysozyme was added to a concentration of 100 μg/mL followed by the addition of Triton X-100 to a concentration of 0.1%. The suspension was incubated at 30 °C for 30 min, sonicated to shear the DNA, and then centrifuged at 12000*g* for 15 min at 4 °C. The insoluble protein was washed once with 50 mM Tris-HCl, pH 8.0, containing 0.1% Triton X-100. The insoluble protein fraction was then dissolved in 6 M guanidine hydrochloride in binding buffer (20 mM Tris-HCl, pH 7.9, 0.5 M NaCl, 5 mM imidazole) and passed over a His-bind resin column (Novagen) according to the manufacturer's instructions, washed with wash buffer (20 mM Tris-HCl, pH 7.9, 0.5 M NaCl, 20 mM imidazole) containing 6 M guanidine hydrochloride (GdnHCl). The proteins were finally eluted with elution buffer (20 mM Tris-HCl, pH 7.9, 0.5 M NaCl, 1 M imidazole) containing 6 M GdnHCl. The eluted protein was next dialyzed against 20 mM Tris-HCl, pH 8.0, to remove the guanidine. Alternatively, the level of GdnHCl in the binding buffer was gradually lowered stepwise to zero (6, 5, 4, 3, 2, 1.5, 1, 0.5, 0.1, 0 M) while the proteins were still bound to the column. The column was washed exten-

sively with binding buffer and wash buffer containing no guanidine, and finally eluted with elution buffer containing no guanidine. The eluted C-propeptides were either used immediately or aliquoted and stored at −20 °C.

Circular Dichroism (CD) Spectra. Samples of C-pro α1 and C-pro α2 were prepared in 10 mM phosphate buffer, pH 7.4, at a concentration of 0.3 mg/mL. The CD spectra were obtained at 25 °C using a JASCO 715 spectropolarimeter (Jacso Corp., Japan) with a quartz cell of 0.1 cm path length (Hellma Cells, Forest Hills, NY).

Molecular Modeling of C-Pro α1 and C-Pro α2. The structures of the human C-pro α1 and C-pro α2 chains of human collagen type I were built from the published sequences on a Silicon Graphics Indigo 2 R4000 X/Z system using Molecular Simulations Insight, Discover 95.0/3.0, and Biopolymer software (Biosym/MSI, San Diego, CA). The starting point for computations was the trans extended chain form of the propeptides. In each case, the two known intrachain disulfide bonds linking the appropriate cysteines were added as specific structural constraints. Using the Discover software, minimization was begun using the general Consistent Valence Force field (CVFF) (*12*) in a Steepest Descents Gradient followed by a Conjugate Gradient until the rms deviation was below 0.001 in each case. The two propeptides were minimized separately, starting from the same extended chain structure.

Preparation of Labeled C-Pro α1, C-Pro α2, and 5′ or 3′ Deleted C-Pro α1 and C-Pro α2 by Coupled Transcription and Translation. The C-pro α1 and C-pro α2 constructs or the 5′ and 3′ deletion constructs were in vitro translated in the presence of [^{35}S]methionine in a TNT-coupled transcription/translation system (Promega). Briefly, 1 μg of plasmid DNA was added to 50 μL of reaction mixture containing 50% rabbit reticulocyte lysate, T7 polymerase, 20 μM amino acid mixture minus methionine, 40 units of RNasin ribonuclease inhibitor, and 50 μCi of [^{35}S]Met (1.4 Ci/μmol). The reaction mixture was incubated at 30 °C for 120 min. After incubation, 1 μL of the mixture was analyzed for efficiency of translation by SDS−PAGE. Unlabeled constructs were translated in essentially the same way, without the addition of [^{35}S]Met but supplemented with 20 μM amino acid mixture minus leucine. By adding both incomplete mixes (amino acid mixture minus methionine and amino acid mixture minus leucine), a sufficient concentration of all amino acids was provided for efficient translation.

In Vitro Protein−Protein Interaction. (A) Binding to C-Pro α1− and C-Pro α2−Sepharose. Purified $(His)_6$-C-pro α1 and $(His)_6$-C-pro α2 were dissolved in 0.5% SDS and coupled to CNBr-activated Sepharose (Sigma Chemicals) according to the manufacturer's instructions. The amount of protein bound to Sepharose was 1 mg/mL for both C-pro α1 and C-pro α2. Equal amounts (50 μL) of C-pro α1−Sepharose and C-pro α2−Sepharose were denatured by suspension in 6 M GdnHCl and then renatured by gradually bringing down the GdnHCl concentration to zero. Equal amounts of the labeled C-propeptides or labeled N- and C-deleted C-propeptides were added to the conjugated Sepharose and allowed to react for 4 h at 4 °C. The tubes were gently rotated during the incubation. After incubation, the Sepharose beads were collected by centrifugation and washed 6 times with 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 0.1% NP40, followed by one wash with 50 mM Tris-HCl, pH 7.5. The C-propeptides associated with the washed beads were eluted by boiling in SDS-gel loading buffer and analyzed by SDS−PAGE. The resulting gels were dried and proteins visualized by autoradiography overnight at −70 °C using Hyper film (Amersham).

(B) Binding of Labeled C-Pro α1, C-Pro α2, and 5′ or 3′ Deleted C-Pro α1 and C-Pro α2 to $(His)_6$-C-Pro α1 and $(His)_6$-C-Pro α2 Immobilized to Ni^{2+}-Sepharose. Cold $(His)_6$-C-pro α1 and $(His)_6$-C-pro α2 were synthesized in rabbit reticulocyte lysates, using the TNT-coupled transcription/translation system (Promega). These mixtures were then added separately to equal amounts of in vitro translated [^{35}S]Met-labeled C-pro α1, C-pro α2, and N- or C-deleted C-pro α1 and C-pro α2 at 37 °C. The reaction temperature was cooled to 4 °C and the left overnight at 4 °C. The next day, 50 μL of His-bind resin (Novagen) was added to the reaction mixture and incubation carried on for another 4 h at 4 °C. The tubes were gently rotated during the incubation. The beads were collected in a microfuge and washed twice with binding buffer and 4 times with wash buffer. The bound C-propeptides were then released by boiling in SDS-gel loading buffer and analyzed by SDS−PAGE followed by autoradiography.

In Vivo Protein−Protein Interaction: Yeast Two-Hybrid Analysis. Fusion proteins of the C-pro α1 and C-pro α2 peptide with the GAL4 DNA binding domain and the Gal 4 activation domain were constructed using the primer pairs OL11, OL8 for C-pro α1 and OL12, OL10 for C-pro α2. The underlined sequences denote the *Nco*I and *Bam*HI restriction sites. The primers were used to amplify the C-pro α1 and C-pro α2 peptide from pHHα1 and pHHα2 as before. The amplified products were then digested with *Nco*I and *Bam*HI and cloned into the *Nco*I and *Bam*HI sites of the yeast two-hybrid GAL4 DNA binding vector pAS2-1 and the GAL4 activation domain vector pACT2 to generate the plasmids pAS2-1-α1 (D1), pAS2-1-α2 (D2), pACT2-α1 (A1), and pACT2-α2 (A2), respectively. Recombinant plasmids were then transferred either individually or simultaneously into the yeast strain CG 1945. The interactions between the various constructs were visually characterized by 5-bromo-4-chloro-3-indoyl β-D-galactopyranoside filter lift assay performed after 3 days of growth on selective medium (*13*). To provide quantitation of the interaction strength, multiple colonies from yeast harboring C-pro α1 and C-pro α2 combinations were grown overnight in liquid medium. Fresh liquid cultures were started and grown for 5 h and assayed for lacZ expression using *o*-nitrophenyl β-D-galactopyranoside.

Production, Purification, and Characterization of Antibodies against C-Pro α1 and C-Pro α2. To prepare antibodies against C-pro α1 and C-pro α2, the $(His)_6$-tagged C-propeptides were expressed and purified on a Ni^{2+}-Sepharose column as described above. The pure propeptides were then used as antigens. The subsequent preparation of the antibodies in rabbits was carried out by Bethyl Laboratories Inc. (Montgomery, TX) according to established procedures. The bacterial strain BL21(DE3) used to express the proteins was solubilized in bicarbonate buffer, pH 8.0, containing 0.1% Triton X-100. The soluble proteins were linked to CNBr-activated Sepharose, to prepare a bacterial proteins−Sepharose column. The antisera against C-pro α1 and C-pro α2 were first passed over this bacterial proteins−Sepharose

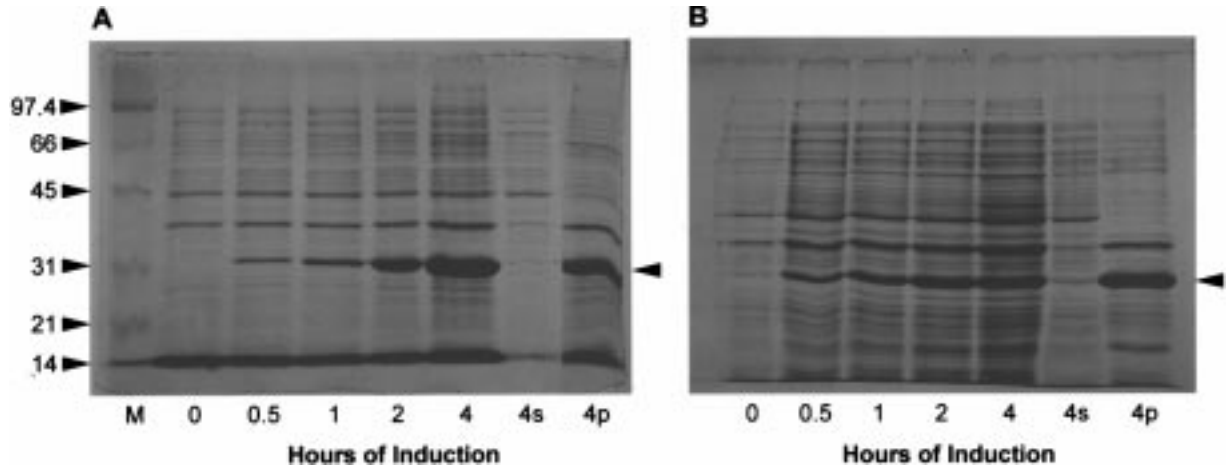

FIGURE 1: Time course of induction of synthesis of recombinant C-pro α1 and C-pro α2, as revealed by SDS-gel electrophoresis analysis. Bacteria expressing recombinant C-pro α1 (A) and C-pro α2 (B) were grown and induced with 1 mM IPTG as described under Materials and Methods. Aliquots at various times of induction were removed and subjected to electrophoresis on a 12% SDS gel. At 4 h, the cells were lysed and separated into a soluble (4s) and an insoluble (4p) fraction. The arrow indicates the induced propeptides.

column to remove antibodies that cross-reacted against the host strain. The C-pro α1 antisera were then passed over a C-pro α2−Sepharose column, to remove antibodies that cross-reacted with C-pro α2. The unbound fractions, now specific to C-pro α1, were collected. Antibodies specific to C-pro α2 were purified in a similar manner with the final exclusion on a C-pro α1−Sepharose column. Antibodies were characterized by Western blot analysis (*12*).

Cell Culture, Metabolic Labeling, and Immunoprecipitation. Human newborn HS68 fibroblasts (ATCC CRL-1635) were grown in DME, supplemented with 10% fetal calf serum, 100 units/mL penicillin, 100 mg/mL streptomycin, and 2.5 μg/mL amphotericin, at 37 °C in a humidified atmosphere of 95% air, 5% CO_2. Prior to labeling, cells were maintained in methionine-free DME medium, supplemented with 10% fetal calf serum, 100 units/mL penicillin, 100 mg/mL streptomycin, and 2.5 μg/mL amphotericin, for 2 h and then metabolically labeled at 37 °C with 20 μCi/mL [^{35}S]Met, for 24 h. At the end of the labeling period, the cell monolayers were rinsed twice with ice-cold PBS and extracted in situ by addition of 4 mL of immunoprecipitation (IP) buffer (0.5% Triton X-100, 0.5% sodium deoxycholate, 150 mM NaCl, 20 mM Tris-HCl, pH 7.4, 1 μg/mL, trypsin inhibitor, 1 μg/mL pepstatin, 1 μg/mL leupeptin, and 0.3 mM phenylmethylsulfonyl fluoride) to the culture flasks. After a 15 min incubation at 4 °C, solubilized components were removed from the monolayer, and 4 mL of fresh IP buffer was added for an additional 15 min. At the end of the second incubation, the two extracts were combined, divided into aliquots of 1.0 mL, and centrifuged in a microcentrifuge at 14 000 rpm for 30 min at 4 °C. Nonspecific binding proteins were first removed from the supernatants by the addition of 50 μL of preimmune serum for 4 h followed by the addition of 50 μL of a 1:1 slurry of protein G−Sepharose (Sigma) and a further incubation of 1 h. The nonspecific immune complexes were then collected by centrifugation at 14 000 rpm for 5 min. The clarified supernatants were immunoprecipitated overnight with either 50 μL of rabbit anti-C-pro α1 or 50 μL of rabbit anti-C-pro α2 antibodies. Following overnight incubations with the antibodies, 50 μL of a 1:1 slurry of protein G−Sepharose (Sigma) was added and incubation carried on for an additional 1 h. The specific immune complexes were sedimented by centrifugation at 14 000 rpm for 10 min. Immunoprecipitates were washed 4 times with 500 μL of IP buffer, once with 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, and finally once with 10 mM Tris-HCl, pH 7.5. Immunoprecipitates were then denatured by boiling for 5 min in SDS sample buffer in the presence of β-mercaptoethanol, and analyzed on 10% SDS−PAGE. Gels were dried and autoradiographed at −70 °C overnight.

RESULTS

Expression, Purification, and Disulfide Bonding of C-Pro α1 and C-Pro α2. The transfected BL21 cells capable of expressing either $(His)_6$-C-pro α1 or $(His)_6$-C-pro α2 were induced with 1 mM IPTG, as described under Materials and Methods, and aliquots were taken at various times of induction. Figure 1 shows that induction of C-pro α1 (Figure 1A) and C-pro α2 (Figure 1B) could be detected as early as 30 min after addition of IPTG. Expression of the recombinant proteins continued to increase until 4 h, the last time point tested. At 4 h, the cell pellets were separated into fractions that were soluble and insoluble in lysis buffer. As is evident from Figure 1A,B, most (>90%) of each recombinant protein was in the insoluble fraction, probably associated with the inclusion bodies. The recombinant C-pro α1 and C-pro α2 in the respective insoluble fractions was dissolved in 6 M GdnHCl and purified to near-homogeneity on a Ni^{2+}-His.Bind resin column. The purified propeptides had slightly different migration rates on SDS-gels (Figure 2, lanes 1, 3), with the C-pro α2 migrating more rapidly although it contained an additional 12 amino acids than C-pro α1. Since the gels were run under denaturing and reducing conditions, the different migration rates could not be attributed to internal disulfide bonding differences, suggesting that the two propeptides might have different conformations.

Dialysis of the eluted proteins in 6 M GdnHCl against 20 mM Tris HCl, pH 8.0, resulted in aggregation and reprecipitation of the recombinant proteins. In an alternate procedure to permit controlled refolding of the recombinant propeptides, the level of GdnHCl in the binding buffer was stepwise lowered to zero while the proteins were still bound to the column. Finally, the column was washed with binding buffer and wash buffer containing no GdnHCl. Then the

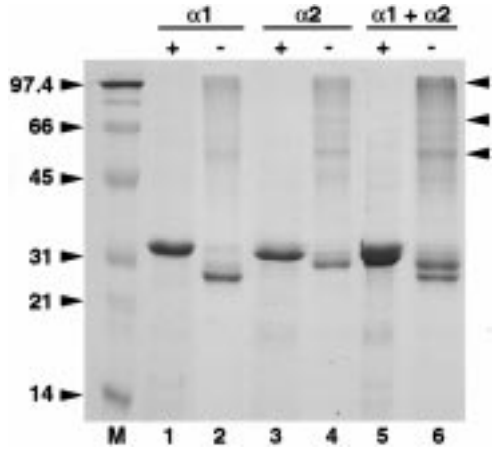

FIGURE 2: Involvement of disulfide bonds in propeptide aggregation. The recombinant C-pro α1 and C-pro α2 were purified and eluted from a Ni^{2+}-His.Bind column. The purified C-pro α1 (lanes 1, 2), C-pro α2 (lanes 3, 4), or a mixture of the α1 and α2 C-propeptides (lanes 5, 6) were allowed to aggregate on dialysis. The aggregates were then dissolved in Laemmli's buffer and boiled either in the presence (+) or in the absence (−) of β-mercaptoethanol and subjected to SDS-gel electrophoresis. Arrows indicate the positions of minor amounts of higher molecular weight complexes when the mixture of α1 and α2 C-propeptides was allowed to aggregate for 48 h. Note the substantially more rapid migration of the intramolecularly disulfide-linked forms of the propeptides. Lane M shows the molecular weight markers.

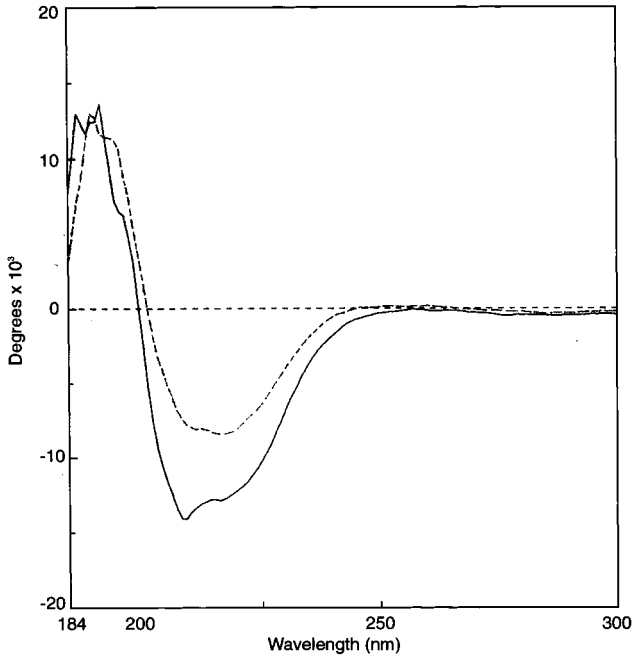

FIGURE 3: Circular dichroism spectra. CD spectra were obtained at 25 °C using a JASCO 715 Spectropolarimeter with a quartz cell of 0.1 cm path length. C-pro α1 (dashed line) and C-pro α2 (solid line) were prepared in 10 mM phosphate buffer, pH 7.4, with both at a concentration of 0.3 mg/mL.

column was eluted with elution buffer containing no Gdn-HCl. Under these conditions, the C-pro α1 and C-pro α2 were eluted in soluble form. However, if the solutions were allowed to stand at 4 °C for 48 h or more, the recombinant proteins aggregated.

It was of interest to see whether the C-propeptide self-aggregates were a result of disulfide bond formation. The aggregates were boiled in Lamelli's buffer, either in the presence or in the absence of β-mercaptoethanol, and subjected to SDS−PAGE. As shown in Figure 2, lanes 1−4, both C-propeptides migrated primarily as monomers regardless of the presence or absence of β-mercaptoethanol. These data suggest that disulfide bond formation did not play a role in their intermolecular association. Nevertheless, in the absence of β-mercaptoethanol, the monomer bands of the C-pro α1 and C-pro α2 migrated faster than when in the presence of β-mercaptoethanol (compare lanes 1 and 3 with lanes 2 and 4, respectively). Thus, the recombinant C-propeptides allowed to refold in monomer form by stepwise reduction of the GdnHCl concentration appear to have formed internal disulfide bonds. The mobilities of the C-pro α1 and C-pro α2 monomer bands, determined in the absence of mercaptoethanol, were different, indicating that the folded internally disulfide-bonded structures of C-pro α1 and C-pro α2 were conformationally distinct.

When monomeric, soluble C-pro α1 and C-pro α2 solutions were combined in equal volumes and concentrations and then allowed to aggregate (Figure 2, lanes 5, 6), the majority of each peptide ran as a monomer on the gels in the absence of β-mercaptoethanol. However, some dimer and trimer formation (arrowheads, Figure 2, lane 6) was evident. These results suggested that the interactions leading to the aggregation of the bacterial synthesized recombinant C-propeptides either with themselves or with each other were not mediated by intermolecular disulfide bond formation. However, after interaction, in the absence of mercaptoethanol, intermolecular disulfide-bonded aggregates could form.

Circular Dichroism. The above data indicated that the monomeric recombinant propeptides formed internal disulfide bonds and produced structures of somewhat different conformations. The conformations of the refolded monomeric propeptides were therefore examined by circular dichrosim in a 10 mM phosphate buffer at pH 7.4. Identical concentrations of 0.3 mg/mL were used. The CD spectra (Figure 3) show that both propeptides have mixtures of β-sheet and α-helix-like conformations, although the values of dichroic ratio were not large. The C-pro α2 propeptide had a strongly more prominent minimum at 210°, showing that it had more β-sheetlike conformation than the C-pro α1.

Molecular Modeling of C-Pro α1 and C-Pro α2. The 2 propeptides are different in length by only 1 amino acid out of a total of 246 for C-pro α1. Comparing the aligned sequences (Figure 4), there is 60.6% identity, and conservative homology between an additional 21.7%. The remaining 17.6% of residues, however, represent nonconservative differences. Of special interest from the structural perspective, five Pro residues in each chain occupy unique positions. No data are available on the structures of the propeptides, but based on the information presented above that the two might have different conformations or packing, we decided to model their conformations by an energy minimization computation. The sequences were entered into the Discover program using the very general CVFF algorithms. The initial sequence entry had the backbone chain in the trans-extended chain conformation, with only two explicit constraints. In C-pro α1, disulfide bonds C81→C244 and C152→C197 were required. The corresponding bonds in C-pro α2 were C84→C245 and C153→C198. No other restraints were used. The first hundred iterations were by the Steepest Descent method; subsequent iterations were by the Conjugate Gradi-

```
α1  D D . . . A N V V R D R D L E V D T T L K S L S Q Q I  24
α2  D Q P R S A P S L R P K D Y E V D A T L K S L N N Q I  27

                                ☆
α1  E N I R S P E G S R K N P A R T C R D L K M C H S D W  51
α2  E T L L T P E G S R K N P A R T C R D L R L S H P E W  54

α1  K S G E Y W I D P N Q G C N L D A I K V F C N M E T G  78
α2  S S G Y Y W I D P N Q G C T M E A I K V Y C D F P T G  81

α1  E T C V Y P T Q P S V A Q K N W Y I S K N P K D K R H 105
α2  E T C I R A Q P E N I P A K N W Y R S S . . K D K K H 106

α1  V W F G E S M T D G F Q F E Y G G Q G S D P A D V A I 132
α2  V W L G E T I N A G S Q F E Y N V E G V T S K E K A T 133

α1  Q L T F L R L M S T E A S Q N I T Y H C K N S V A Y M 159
α2  Q L A F M R L L A N Y A S Q N I T Y H C K N S I A Y M 160

α1  D Q Q T G N L K K A L L L K G S N E I E I R A E G N S 186

α2  D E E T G N L K K A V I L Q G S N D V E L V A E G N S 187

                                ☆
α1  R F T Y S V T V D G C T S H T G A W G K T V I E Y K T 213
α2  R F T Y T V T V D G C S K K T N E W G K T I I E Y K T 214

α1  T K T S R L P I I D V A P L D V G A P D Q E F G F D V 240
α2  N K P S R L P F L D I A P L D I G G A D H E F F V D I 241

α1  G P V C F L                                         246
α2  G P V C F K                                         247
```

FIGURE 4: Comparison and alignment of the sequences of the C-pro α1 and C-pro α2 peptides. The sequences of the human C-propeptides are shown and numbered from the N-terminal residue in the propeptide. The asterisks denote the first and last amino acids in the N-terminal (5′) and C-terminal (3′) truncated constructs, respectively. The differences in the placement of conformationally significant proline residues have been highlighted, and the intramolecular disulfide bonds are joined by arrows. The internal disulfide-bonded loop sequence is underlined.

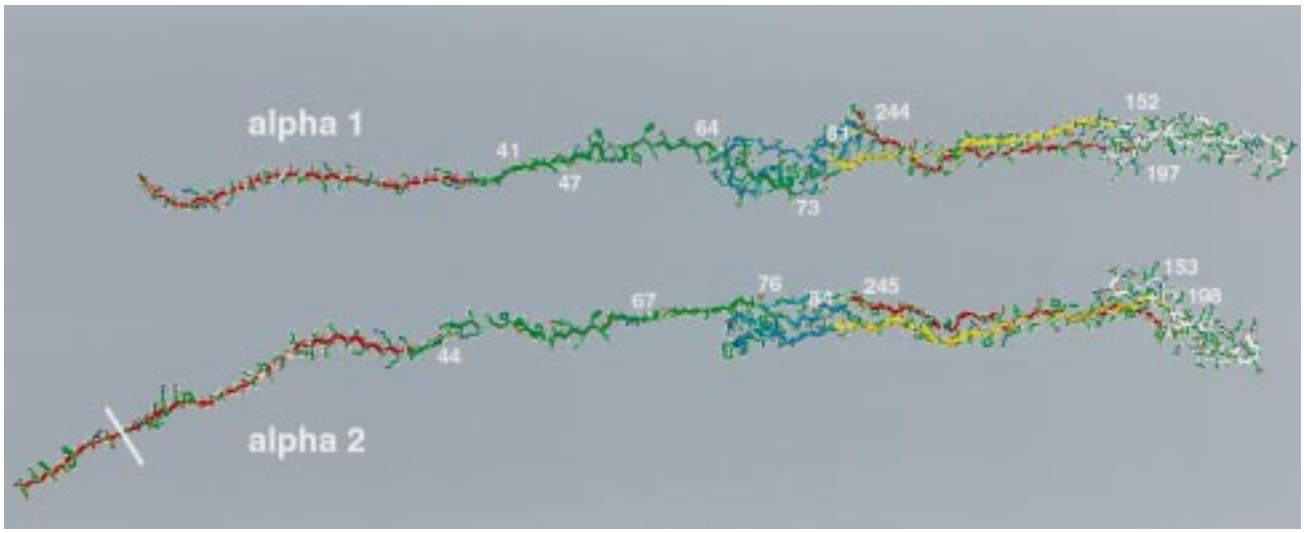

FIGURE 5: Molecular modeling of C-pro α1 and C-pro α2. The structures of C-pro α1 and C-pro α2 were built from published sequences in an extended form on a Silicon Graphics Indigo 2R4000 X/Z graphics. The extended chain structures were constrained by the formation of the two natural disulfide bonds in each case. The deduced minimum energy three-dimensional structure of each peptide is shown. The domains with backbones in red show the portions removed in the 5′ and 3′ deletion constructs. The remaining initial sequence (green) continues as an extended chain until the first globular domain, G-I (blue), is formed. The polypeptide chain leaves G-I as an extended chain (yellow) until the G-II domain forms (white). Finally, the chain leaves G-II to loop back to form the C81→C244 or C84→C245 disulfide bond. The antiparallel chains in the L domain show less structure in C-pro α1 than in C-pro α2. The white line transecting the C-pro α2 shows the telopeptide−propeptide junction. The numbering system used in describing the propeptides in Figure 4 begins at this point. The numbers in this figure show the positions of the Cys residues.

ent method. Minimization was carried on until the rms deviation between conformations was less than 0.001.

The overall conformation for C-pro α1 converged slowly to a net total energy of −355 kcal/mol, and 40 000 iterations were required. Surprisingly, the C-pro α2 conformation, using identical protocols, converged more rapidly (10 000 iterations) to a total energy of −187 kcal/mol. The models (Figure 5), although quite similar, are nevertheless different. Both begin with an N-terminal domain that is essentially an extended chain. This region contains the first two (α2, C44,C67) or three (α1, C41,47,64) Cys residues that are not involved in intramolecular disulfide bonding (the numbering is according to Figure 4). In each case, a globular domain (G-I) of three antiparallel chains then follows (green, blue

in Figure 5), folded so that C81 (α1) and C83 (α2) are at the end of a loop disulfide-linked to C244 or C245, respectively. The peptide chain continues as an extended chain (yellow) until a second globular domain (G-II) (white), containing the second small disulfide-bonded loop (α1, C152→C197, α2, C153→C198), is formed. The C-terminal sequence then continues as an extended chain (red) but looped back to permit formation of the final C81→C244 and C82→C245 disulfide bonds (Figure 5). The two antiparallel extended chains form a linkage (L) domain between the G-I and G-II in both C-propeptides. The L domains may be flexible and act as hinge regions during C-propeptide interactions. The differences in the α1 and α2 G-I and G-II globular domains, evident in Figure 5, may be important in setting the specificity of chain selection interactions.

In Vitro Protein−Protein Interaction. Before studying the interactions of the recombinant C-propeptides, it was necessary to verify that in vitro synthesized C-propeptides were also capable of folding correctly. [^{35}S]Met-labeled C-propeptides were synthesized in rabbit reticulocyte lysate using the TNT-coupled transcription/translation system. The labeled translation products were boiled in electrophoresis buffer (*11*), either in the presence or in the absence of *β*-mercaptoethanol, and subjected to SDS−PAGE (Figure 6A). In the presence of *β*-mercaptoethanol, nearly all the in vitro synthesized protein ran as a monomer, with very little present as dimer. However, in the absence of *β*-mercaptoethanol, there was an increase in the dimeric form of both propeptides with a corresponding reduction in the monomer content. Moreover, the monomeric forms ran as doublet bands, suggesting, as in the case of the recombinant propeptides, the formation of intramolecular disulfide bonds and a more folded conformation. Again, there was a difference in the migration rates between the C-pro α1 and C-pro α2 propeptides.

The first approach to quantification of the interactions of the C-propeptides with themselves and with each other was to test whether [^{35}S]Met-labeled C-propeptides, produced in the coupled transcription−translation system, could be retained on a C-pro α1−Sepharose or a C-pro α2−Sepharose affinity column. The C-pro α1−Sepharose and C-pro α2−Sepharose affinity columns were prepared by immobilizing bacterial synthesized monomeric recombinant C-pro α1 and C-pro α2 to CNBr-activated Sepharose in the presence of SDS.

The C-pro α1−Sepharose affinity column was capable of binding both in vitro synthesized [^{35}S]Met-labeled C-pro α1 and C-pro α2 (Figure 6B, lanes 1 and 2). Identical amounts of labeled proteins were applied in each case. Surprisingly, in addition to binding C-pro α1, the C-pro α2-affinity column was also capable of binding in vitro synthesized [^{35}S]Met-labeled C-pro α2 to an equivalent extent (Figure 6B, lanes 3 and 4).

The preceding experiments were done on the C-propeptides synthesized in bacteria and immobilized to Sepharose under conditions in which the majority of the propeptides were in the monomeric form. Since propeptides synthesized in the rabbit reticulocyte lysates showed the presence of monomer folding and some dimer formation, we decided to study the interactions entirely with the C-propeptides synthesized by rabbit reticulocyte lysates. Briefly, cold $(His)_6$-C-pro α1 or $(His)_6$-C-pro α2 was synthesized in rabbit reticulocyte lysates and added to [^{35}S]Met-labeled C-pro α1 or C-pro α2, lacking the $(His)_6$-tag. Identical amounts of labeled proteins were applied in each case. The reactions were allowed to continue overnight, following which His-bind resin was used to capture the $(His)_6$-propeptides. Any radioactivity brought down with the His-bind resin would be due to interaction of the labeled peptides with the unlabeled $(His)_6$-C-propeptides. The results are shown in Figure 6D. As before, C-pro α1 and C-pro α2 were both capable of binding to themselves and to each other (lanes 1, 4, 7, 10). The interactions were thus comparable between the recombinant and in vitro synthesized constructs. The amount of initial radioactive propeptide added to the mixture is shown in Figure 6C. Thus, comparing the intensities of labeling in panels C and D, lanes 1, 4, 7, and 10, C-pro α1 and C-pro α2 appear to bind to each other and to themselves with similar efficiencies.

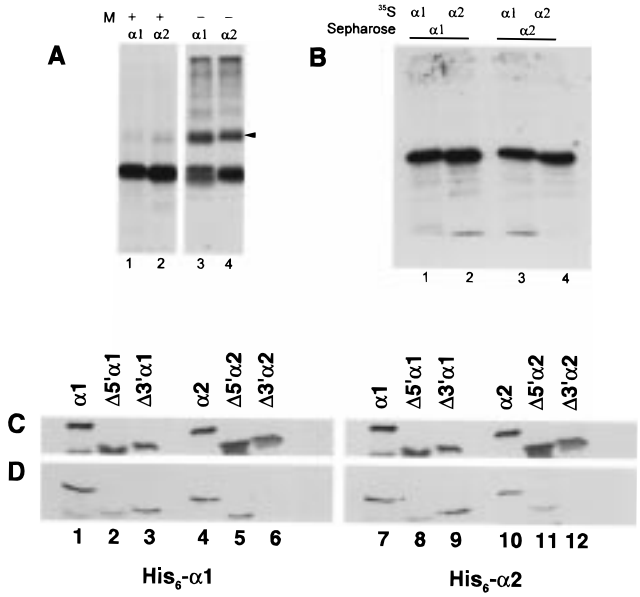

FIGURE 6: In vitro protein−protein interaction. (A) Formation of disulfide bonds. Rabbit reticulocyte lysate synthesized [^{35}S]Met-labeled C-pro α1 (lanes 1 and 3) and C-pro α2 (lanes 2 and 4) were subjected to SDS−PAGE either in the presence (lanes 1, 2) or in the absence (lanes 3, 4) of *β*-mercaptoethanol. Following electrophoresis, the gel was fixed, dried, and analyzed by autoradiography overnight. The arrow indicates the position of dimers formed in the absence of *β*-mercaptoethanol. (B) Sepharose-linked C-pro α1− (lanes 1, 2) or C-pro α2−Sepharose (lanes 3, 4) affinity columns were reacted with equal amounts of in vitro synthesized [^{35}S]Met-labeled C-pro α1 (lanes 1, 3) or C-pro α2 (lanes 2, 4). The beads were washed, and the bound radioactive peptides were eluted by boiling in Laemmli's buffer in the presence of *β*-mercaptoethanol and subjected to SDS-gel electrophoresis. The C-pro α2 interaction with itself (lane 4) is essentially as strong as the C-pro α1−C-pro α1 and C-pro α2−C-pro α1 interactions. (C, D). Equal amounts of radioactive in vitro translated C-propeptides and deletion propeptides were taken for binding to unlabeled $(His)_6$-C-pro α1 (lanes 1−6) and $(His)_6$-C-pro α2 (lanes 7−12). Panels C represent the amount of radioactivity taken for binding. Panels D show the amounts of in vitro synthesized [^{35}S]Met-labeled C-pro α1 (lanes 1, 7), C-pro α2 (lanes 4, 10), Δ5′C-pro α1 (lanes 2, 8), Δ3′C-pro α1 (lanes 3, 9), Δ5′C-pro α2 (lanes 5, 11), or Δ3′C-pro α2 (lanes 6, 12) retained by the chelation-linked $(His)_6$-C-pro α1 or $(His)_6$-C-pro α2. The Novogen His.Bind beads were pelleted and washed. The bound radioactively labeled peptides were eluted by boiling in Laemmli's buffer in the presence of *β*-mercaptoethanol and subjected to SDS-gel electrophoresis. The gels were fixed, dried, and analyzed by autoradiography overnight.

Table 1: Intensity of Binding between Combinations of C-Pro α1 and C-Pro α2 in the Yeast Two-Hybrid System

interacting propeptides	% β-galactosidase act. ± SD[a]
Dα1Aα1	100 ± 2.5
Dα1Aα2	98 ± 1.5
Dα2Aα1	92 ± 6
Dα2Aα2	67.5 ± 1.6

[a] Average of 2 separate experiments, 12 individual readings for each experiment. The intensity of binding between C-pro α1 with itself was arbitrarily taken as 100%.

This approach also provided a convenient means to examine the specificity of the interactions in terms of the propeptide domains. The 5′ and 3′ deletion constructs of C-pro α1 and C-pro α2 were prepared. The deleted sequences can be seen in Figure 5 as the chain backbone segments indicated in red. The 5′ deletions exclude the N-terminal portions D1 to P37 in C-pro α1, and D1 to P40 in C-pro α2. All of the Cys residues are retained. As shown in Figure 6D, lanes 2 and 8, deletion of the N-terminal region of C-pro α1 substantially diminished its binding to either itself or C-pro α2. Similarly, deletion of the N-terminal end of C-pro α2 also markedly diminished binding to itself or to C-pro α1 (lanes 5, 11). Deletion of the C-terminal end of C-pro α1, on the other hand, had only a small effect on binding to either propeptide (lanes 3, 9). The C-terminal deleted C-pro α1 bound equally well to C-pro α1 or C-pro α2. In marked contrast, deletion of the C-terminal segment of C-pro α2 completely abolished its binding to both C-pro α1 and C-pro α2 (lanes 6, 12). Similar results were obtained for binding of the N- and C-terminal deletions of the C-pro α1 or C-pro α2 constructs, to CNBr-linked C-pro α1 and C-pro α2 (results not shown). The fact that similar deletion constructs of the C-pro α1 and C-pro α2 yielded such different binding results supports the conclusion that the interactions being studied are very specific in nature.

In Vivo Protein−Protein Interaction. Yeast Two-Hybrid Experiments. The in vitro experiments described above used either purified recombinant propeptides or propeptides synthesized and folded in cell-free systems. In normal cellular synthesis, it is possible that the local environment and the presence of other cell proteins could modulate the folding of the propeptides and/or the propeptide interactions. To approach the in vivo system, the yeast two-hybrid system was studied. Fusion proteins were constructed in which the DNA binding domain of Gal4 was fused to either C-pro α1 or C-pro α2 (D1, D2). Similarly, the activation domain of Gal4 was fused to C-pro α1 and C-pro α2 (A1, A2). Interactions were tested by cotransformation of different combinations of the DNA binding domain fusion proteins (D) with the activation domain fusion proteins (A), into the yeast strain CG 1945. This strain harbors an integrated lacZ reporter gene driven by the Gal1 upstream activating sequence (UAS) and three tandem copies of the Gal4 17-mer consensus sequence. In this system, interactions between the C-propeptide portions of the D and A constructs are required to drive the expression of the lacZ reporter gene.

The relative strengths of interactions between the C-propeptides were assessed both by a direct lacZ plate assay and by quantitation of the reporter activity in a liquid culture (Table 1). The lacZ plate assay (Figure 7) showed unequivocally that, in vivo, C-pro α1 interacted with both itself and

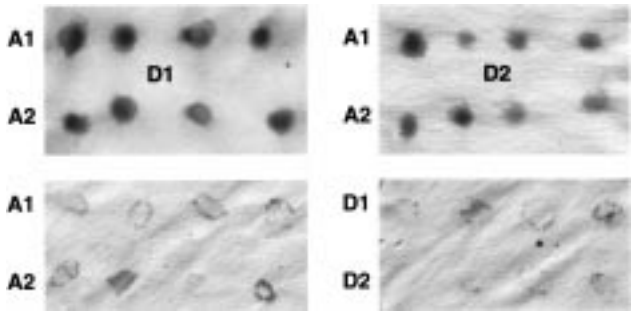

FIGURE 7: Interactions between C-pro α and C-pro α2 in the yeast two-hybrid system. Four different individual yeast colonies containing combinations of the DNA binding domain and activation domain fusion proteins (D1A1, D1A2, D2A1, and D2A2) were grown on SD/−Trp/−Leu/ for 3 days. As controls, yeast cells containing fusion proteins with the activation domain alone (A1, A2) or the DNA binding domain alone (D1, D2) were plated on SD/−Leu or SD/−Trp, respectively. After growth for 3 days, the colonies were assayed for the expression of β-galactosidase by the colony lift method. Colonies that contained both plasmids turned blue while those containing the individual plasmids alone stayed pink.

C-pro α2. Likewise, the C-pro α2 interacted with C-pro α1 and itself, confirming that the interaction of C-pro α2 with itself is real and not an artifact of the in vitro system. However, quantitation of the interactions in the liquid assay showed that in the yeast cells interaction of C-pro α2 with itself was significantly weaker than the C-pro α1−C-pro α1 interaction, whereas the C-pro α1−C-pro α2 interaction was equivalent to the C-pro α1−C-pro α1 interaction (Table 1).

In Vivo Protein−Protein Interactions. Western Blot and Immunoprecipitation Analysis. Since the C-pro α2 → C-pro α2 interaction was significantly weaker than the C-pro α1 → C-pro α2 interaction in the quantitative yeast two-hybrid system assay, whereas their interactions were more closely equivalent in vitro, it seemed logical to determine if there were any other cell proteins that might specifically bind to C-pro α2 and thus prevent self-association. For this investigation, antibodies were raised against the recombinant C-pro α1 and C-pro α2 and carefully purified, as described under Materials and Methods, to minimize cross-reactivity. Whole uninduced and IPTG-induced bacterial cell extracts were used for Western blot analysis. The blots, shown in Figure 8, demonstrate that the purified antibodies raised against C-pro α1 were specific to α1 (panel A, lanes 2 and 4), and that those against C-pro α2 were specific to α2 (panel B, lanes 2 and 4) without appreciable cross-reactivity. This finding was confirmed by ELISA (data not shown). The anti-C-pro α1 interaction with C-pro α2 (Figure 8A, lane 4) was minor. There was no interaction of the anti-C pro α1 with the bacterial proteins whether C-pro α1 was present or absent.

The interaction of anti-C-pro α2 with C-pro α1 was also minor (Figure 8B, lane 2). However, as shown in Figure 8B, lane 4, when C-pro α2 was present in large amount, the anti-C-pro α2 antibody reacted with a number of additional proteins. The lower molecular weight bands seen in lane 4 (Figure 8B) could be due to either incomplete synthesis and/or degradation products of the induced C-pro α2. The higher molecular weight bands could arise from the presence of stable aggregates of C-pro α2 with the cell proteins. In any case, the additional bands seen in lane 4 cannot be attributed to nonspecific background reaction with bacterial proteins, as the same amount of bacterial protein was present in all

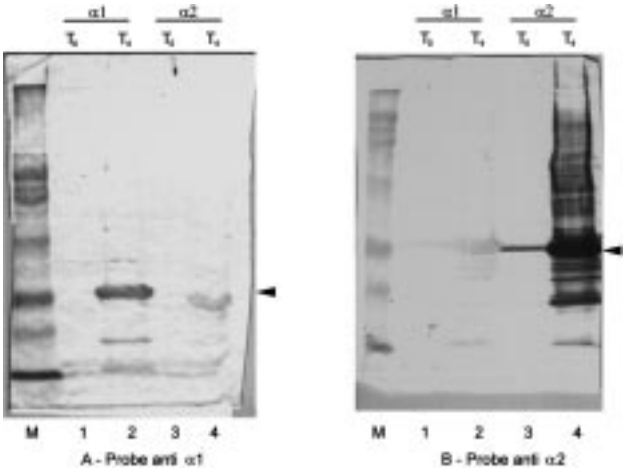

FIGURE 8: Demonstration of specificity of the antibodies to the propeptides. Cell extracts of bacteria expressing C-pro α1 (lanes 1 and 2) or C-pro α2 (lanes 3 and 4) either before induction (T_0) or 4 h after induction (T_4) were run on a 12% SDS gel and transferred to nitrocellulose. Lane M shows the prestained molecular weight markers. The blots were blocked with BSA and reacted overnight at 4 °C with the antibodies that had been prepared against the recombinant C-pro α1 (A) or C-pro α2 (B). The antigen−antibody complexes were visualized by reaction with anti-rabbit IgG-linked alkaline phosphatase and the substrate BCIP/NBT.

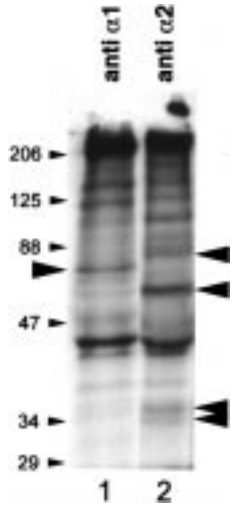

FIGURE 9: Possible role of accessory proteins in regulating in vivo C-propeptide interactions. [^{35}S]Met metabolically labeled proteins of a human fibroblast, HS68, cell line total cell lysate were first interacted with preimmune serum and protein G−Sepharose. The clarified supernatants were then subjected to immunoprecipitation with either anti-C-pro α1 (lane 1) or anti-C-pro α2 (lane 2). The immune complexes were brought down with protein G−Sepharose. The pellets were washed as described under Materials and Methods and then subjected to SDS-gel electrophoresis followed by autoradiography. The molecular masses of the standards are shown on the left by the small arrows. The large arrows show the uniquely immunoprecipitated proteins in each case.

four lanes of Figure 8B. Lanes 1, 2, and 3 showed no reaction of the anti-C-pro α2 with bacterial proteins in the absence of excess C-pro α2. The band of antibody labeled C-pro α2 seen in Figure 8, lane 3 suggests that the control of C-pro α2 synthesis is leaky, with a low level of synthesis even in the absence of ITPG induction. This is also evident in Figure 1B, lane 0. In the presence of this small amount of C-pro α2, no background staining of other proteins was detected. Thus, the data in Figure 8, lane B4 shows that C-pro α2 must interact and bind to several bacterial cell proteins much more strongly than does C-pro α1. The preimmune serum failed to pick up any specific bands (data not shown).

The purified, monospecific antibodies were then used to immunoprecipate reactive proteins from a human fibroblast (HS68) cell lysate, in which the proteins had been metabolically labeled with [^{35}S]Met. As shown in Figure 9 (lanes 1, 2), anti-C-pro α1 and anti-C-pro α2 immunoprecipitated a number of similar proteins. The heavy, high molecular mass band in both lanes is probably related to the completed collagen pro-α chains. However, there were specific differences. The anti-C-pro α1 specifically immunoprecipitated a protein of calculated molecular mass 76 K (left arrowhead), which was not immunoprecipitated with anti-C-pro α2. The anti-C-pro α2, on the other hand, immunoprecipitated several proteins not reactive with anti-C-pro α1. The most prominent anti-C-pro α2-precipitated protein band had molecular mass 65 K (Figure 9, lane 2, right arrowhead). Other proteins specifically coprecipitated by anti-C-pro α2 were at molecular mass 85, 35, and 33 K. Thus, the C-pro α1 and C-pro α2 peptides have distinctly different interaction capabilities with other cell proteins, in both human fibroblasts and bacterial cells.

DISCUSSION

As noted in the introduction, the initial step in type I procollagen heterotrimer formation is the association of the C-propeptides of the α chains. The triple helix is then propagated from the C- to the N-terminus in a zipper-like manner (*1*−*4*). The basic question of the selectivity of heterotrimer formation can be approached from either of two perspectives. What specific interactions drive the selection of the [pro α1]$_2$[pro α2] heterotrimer, or how is the formation of the incorrect [pro α1]$_3$ and [pro α2]$_3$ homotrimers or the [pro α1][pro α2]$_2$ heterotrimer suppressed? The data presented above all demonstrate that purified recombinant C-pro α1 and C-pro α2 both have a tendency to self-aggregate. At this point, the nature of the interacting and stabilizing forces remains uncertain. However, the data shown in Figure 2 demonstrate that interchain disulfide bond formation is not likely to be a prerequisite for interaction and in-register triple-helix formation in the heterotrimer. The SDS-gel electrophoresis profiles of the bacterial synthesized C-propeptides in the presence or absence of β-mercaptoethanol were essentially identical, with most of the protein running as the monomer. The same results were obtained when the purified α1 and α2 C-propeptides were mixed and allowed to aggregate, thus eliminating the formation of intermolecular disulfide bonds as the mechanism for aggregation. This is in keeping with recent results showing that interchain disulfide bonds do not play a role in the interactions between the C-propeptides of type III collagen during molecular assembly (*15*). It has also been shown that type I procollagen chains can associate into a triple-helical molecule even when interchain disulfide bond formation is inhibited (*16*).

Although interchain disulfide bonds were not formed, intrachain disulfide bonds were formed as made evident by the faster migration of C-pro α1 and C-pro α2 in SDS-gels in the absence of β-mercaptoethanol. Thus, we can assume that the C-pro α1 and C-pro α2 were intramolecularly folded. Moreover, in the absence of β-mercaptoethanol, C-pro α1 ran slightly faster than C-pro α2, indicating that the two propeptides might be folded differently.

The folding question is of major importance to the interpretation of the data presented here. There is always the question of how closely proteins produced by recombinant means are folded relative to their native state. Indeed, the recombinant proteins, as produced, were not soluble in the bacterial lysates without the addition of 6.0 M GdnHCl. However, the $(His)_6$-tagged propeptides could be linked as monomer to the His.bind Ni-chelating column in 6.0 M GdnHCl. The stepwise reduction of the GdnHCl concentration permitted folding to take place while the bound propeptides could not interact and they could be eluted as monomeric, intramolecularly folded and disulfide-bonded molecules. Their conformations and interactions were then examined in that state.

The CD spectra obtained on the two propeptides under identical conditions and concentrations, Figure 3, demonstrated unequivocally that the two have different conformations. Analysis of the CD spectra (*17*, *18*) suggests that C-pro α2 and C-pro α1 have different proportions of α-helix, *β*-sheet, and random chain conformations. The models produced by energy minimization (Figure 5) are consistent with these data. The propeptides, with their high degree of sequence identity and homology, have similar overall structures, with two clearly defined folded domains, G-I and G-II. However, the equivalent folded domains of the two propeptides are not identical in structure. The antiparallel chains of the L linkage domains joining G-I to G-II are similar in length but are also not of equivalent structure. We will return to the differences in structures later in this discussion, as the basis for the differing interaction behavior of the propeptide deletion constructs.

Rabbit reticulocyte lysates have been extensively used to check protein−protein interactions (*19*, *20*) and are known to correctly fold the synthesized proteins (*21*, *22*). The C-propeptides synthesized by translation in rabbit reticulocyte lysates also folded and formed intramolecular disulfide bonds. These monomeric, disulfide-bonded propeptides migrated more rapidly on SDS-gels than did their reduced (+mercaptoethanol) forms (lanes 3 and 4 of Figure 6A). This is especially evident for the C-pro α1, where the appearance of two distinct monomer bands is obvious in lane 3A. Thus, one can assume that the reticulocyte lysate-produced propeptides were essentially properly folded. The binding studies were carried out under conditions that would retain this folding of the C-propeptides. Since the binding data for the recombinant propeptides were equivalent to those for the reticulocyte lysate-produced peptides, it is reasonable to conclude that the recombinant proteins, after stepwise renaturation, were also in an essentially native folded conformation. In the case of the CNBr−Sepharose-linked propeptides, the propeptides were linked under conditions in which they were monomeric. Before being used in the binding experiments, the linked peptides were denatured in 6 M GdnHCl and then stepwise renatured. In the case of the His.bind resin pull-down experiments, all of the interacting propeptides and deletion constructs were synthesized in rabbit reticulocyte lysates.

The formation of a small amount of disulfide bond mediated dimer in the reticulocyte lysate produced C-pro α1 + C-pro α2 mixture (Figure 6A) was unmistakable. Nevertheless, the intermolecular interactions were present and strong even where interchain chain disulfide bonding as in the bacterial system was not present. Thus, the formation of the disulfide bonds is not required to drive the interchain association, either for C-pro α1 → C-pro α1, C-pro α1 → C-pro α2, or C-pro α2 → C-pro α2. The determinants of the protein−protein interaction must therefore lie in the amino acid sequences and conformations of the interacting domains. The same conclusion had been reached for type III collagen, in which a discontinuous sequence of 15 amino acids present in the C-propeptide was shown to direct self-association (*23*). The interaction behavior of the peptides derived from the N- and C-terminal deletion constructs as shown in Figure 6D is especially instructive. In each case, as depicted in Figure 4, the central part of each propeptide was retained, with all cysteines intact, except for the most C-terminal Cys residue in the C-terminal deletions. Most studies (*24*) have emphasized the conserved sizes and sequences of the C-propeptides of type I and III procollagens. Indeed, comparing the human C-pro α1 and C-pro α2 sequences, it is evident that the sequence differences are in most cases very conservative. However, as emphasized in Figure 4, the N-terminal sequences that were deleted to produce the Δ5′C-pro α1 and Δ5′C-pro α2 constructs are also quite different. Three Pro residues present in C-pro α2 are not matched in the C-pro α1 sequence. Thus, the two propeptides must have distinctly different conformations in this region. Similarly, in the C-terminal deletion region, the C-terminal sequences of the two propeptides each have one uniquely placed Pro residue that ensures different conformations for the two sequences. The central globular domains also have strategically placed nonconservative sequence differences that clearly make the two propeptides different in their folded conformations.

The binding studies shown in Figure 6D can be interpreted from this perspective. The 36 amino acid deletion from the C-terminal end of C-pro α1 (depicted in red in Figure 5) had little effect on its binding to itself or to C-pro α2 (compare lanes 1 and 3 with lanes 7 and 9, Figure 6). The same deletion in C-pro α2 completely abolished binding to itself and to C-pro α1 (compare lanes 4 and 6 with lanes 10 and 12). These results indicate that the deleted C-terminal portion of the C-pro α2 peptide was essential for binding to itself or to C-pro α1. The structural models in Figure 5 support this by suggesting the importance of the C-terminal L domain of C-pro α2 to its conformation, whereas the equivalent sequence in C-pro α1 is less structured.

The N-terminal peptide deletion constructs provide a somewhat more ambiguous picture. The binding of Δ5′C-pro α1 to intact C-pro α1 and intact C-pro α2 is diminished (Figure 6C, lanes 1, 2 and 7, 8) but not abolished. Removal of the same region of C-pro α2 (Δ5′C-pro α2) also leads to an inhibition of binding to itself and to C-pro α1. The inhibition of binding on removal of the N-terminal end of either C-pro α1 or C-pro α2, together with the complete lack of binding on removal of 36 amino acids from the C-terminus of C-pro α2, strengthens the conclusion that the interactions seen between the intact C-propeptides are specific. The present experimental procedures, however, do not enable us to determine the stoichiometry of binding. The experiments described above cannot specify whether C-pro α1 binds to itself and C-pro α2 as a monomer or as a dimer, nor rule out the possibility that even larger aggregates might exist.

Synthesis of foreign proteins in yeast has been shown to be favorable to proper folding and disulfide bond formation of the recombinant proteins and domains (*25*). The yeast two-hybrid system has also recently been used to show interactions between type IV collagen and type VI collagen (*26*). The yeast two-hybrid system was therefore used here to probe the interaction of the C-propeptides in a near in vivo situation. The data obtained in the yeast two-hybrid system once again showed that C-pro α2 is capable of binding to itself in vivo, although with only 68% of the affinity shown by the self-interaction of C-pro α1. On the other hand, C-pro α1 bound to itself and to C-pro α2 with nearly equal affinities (Table 1). These results are in keeping with a recent report by Tomita et al., demonstrating the presence of pro α2(1) dimers and unstable trimers in a baculovirus expression system (*27*). In view of the fact that C-pro α2 is also capable of binding to itself in vitro, whereas the presence of [pro $\alpha2(1)]_3$ homotrimers has not been detected so far in any in vivo or in vitro system, suppression of the C-pro α2 self-interaction seems to be the more likely possibility. The observation that while the binding of C-pro α2 to itself was strong in vitro but it was weaker in vivo raised the possibility that there might be other proteins in the cell that interact with C-pro α2 and so prevent it from associating with itself in vivo. The yeast cell may contain homologues of those proteins that interact with the C-pro α2 peptide in collagen-producing fibroblasts. The fact that C-pro α2 is over-expressed and the possibility that the yeast homologues do not bind as strongly as their mammalian counterparts could explain why a reduction rather than a complete abolition of binding is seen in the yeast two-hybrid system. The co-immunoprecipitation data (Figure 9), showing that the antibody against C-pro α2 coprecipitates specific proteins that were not immunoprecipitated by the antibody against C-pro α1, strengthens the idea that the C-pro α2 domain reacts with specific and distinct cellular proteins. The interaction of these proteins with the C-propeptide of the α2 chain may be important in preventing assembly of an α2 homotrimer. This notion is similar to the conclusions already reached concerning the role of Hsp 47 in interacting specifically with the N-propeptide of the pro α1 chain preventing its premature interaction (*28*). It is likely that these proteins will be located in either the endoplasmic reticulum or the Golgi compartments of collagen producing cells. The isolation and characterization of the proteins interacting with the C-pro α1 and C-pro α2 chain regions and a detailed study of the interactions of the recombinant propeptides will be the next steps toward understanding the mechanism of chain selection and heterotrimer formation in procollagen I.

REFERENCES

1. Schofield, D. J., Uitto, J., and Prockop, D. J. (1974) *Biochemistry 13*, 1801−1806.
2. Olsen, B. R., Hoffman, H. P., and Prockop, D. J. (1976) *Arch. Biochem. Biophys. 175*, 341−350.
3. Bruckner, P., Bachinger, H. P., Timpl, R., and Enjel, J. (1978) *Eur. J. Biochem. 90*, 595−603.
4. Engel, J., and Prockop, D. J. (1991) *Annu. Rev. Biophys. Biochem. 20*, 137−152.
5. Jimenez, S. A., Bashey, R. I., Benditt, M., and Yankowski, R. (1977) *Biochem. Biophys. Res. Commun. 78*, 1354−1361.
6. Uitto, J. (1979) *Arch. Biochem. Biophys. 192*, 371−379.
7. Hu, G., Tylzanowski, P., Inoue, H., and Veis, A. (1995) *J. Cell. Biochem. 59*, 214−234.
8. Doege, J. K., and Fessler, H. J. (1986) *J. Biol. Chem. 261*, 8924−8935.
9. Bruckner, P., Eikenberry, E. F., and Prockop, D. J. (1981) *Eur. J. Biochem. 118*, 607−613.
10. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
11. Laemmli, U. K. (1970) *Nature 227*, 680−685.
12. Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., and Hagler, A. T. (1988) *Proteins: Struct., Funct., Genet. 4*, 31−47.
13. Bohen, S. P., and Yamamoto, K. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11424−11428.
14. Towbin, H., Staehlin, T., and Gordon, J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350−4354.
15. Bulleid, N. J., Wilson, R., and Lees, J. F. (1996) *Biochem. J. 317*, 195−202.
16. Beck, K., Boswell, B. A., Ridgway, C. C., and Bachinger, H. P. (1996) *J. Biol Chem. 271*(35), 21566−21573.
17. Sreerama, N., and Woody, R. W. (1993) *Anal. Biochem. 209*, 32−44.
18. Sreerama, N., and Woody, R. W. (1994) *Biochemistry 33*, 10022−10025.
19. Yanagiswa, J., Takahashi, M., Kanki, H., Yano-Yanagisawa, H., Tazunoki, T., Sawa, E., Nishitobs, T., Kamishohara, M., Kobayashi, E., Kataoka, S., and Sato, T. (1997) *J. Biol. Chem. 272*, 8539−8545.
20. Hogenesch, J. B., Chan, W. K., Jackiw, V. H., Brown, R. C., Gu, Y., Pray-Grant, M., Perdew, G. H., and Bradfield, C. A. (1997) *J. Biol. Chem. 272*, 8581−8593.
21. Pratt, W. B., and Welsh, M. J. (1994) *Semin. Cell Biol. 5*(2), 83−93.
22. Hutchison, K. A., Scherrer, L. C., Czar, M. J., Stancato, L. F., Chow, Y. H., Jove, R., and Pratt, W. B. (1993) *Ann. N.Y. Acad. Sci. 684*, 35−48.
23. Lees, J. F., Tasab, M., and Bulleid, N. J. (1997) *EMBO J. 16*, 908−916.
24. Ramirez, F. (1989) in *Collagen−Molecular Biology* (Olsen, B. J., and Nimmi, M. E., Eds.) Vol. 4, pp 21−30, CRC Press, Inc., Boca Raton, FL.
25. Handford, P. A., Baron, M., Mayhew, M., Willis, A., Beesley, T., Brownlee, G. G., and Campbell, I. D. (1990) *EMBO J. 9*, 475−480.
26. Kuo, H.-J., Maslen, C. L., Keene, D. L., and Glanville, R. W. (1996) *J. Biol. Chem. 272*, 26522−26529.
27. Tomita, M., Kitajima, T., and Yoshizato K. (1997) *J. Biochem. 121*, 1061−1069.
28. Hu, G., Gura, T., Sabsay, B., Sauk, J., Dixit, S. N., and Veis, A. (1995) *J. Cell. Biochem. 59*, 350−367.

BI9821824