Structure and Assembly of the Heterotrimeric and Homotrimeric C-Propeptides of Type I Collagen: Significance of the α2(I) Chain†

James P. Malone,‡ Keith Alvares, and Arthur Veis*

Department of Cell and Molecular Biology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611

Received May 5, 2005; Revised Manuscript Received September 9, 2005

ABSTRACT: Assembly of the type I procollagen molecule begins with interactions among the C-pro α1(I) and C-pro α2(I) domains. The C-propeptide domains themselves have subdomains of distinct structures. The important questions are where chain association begins and the basis of the chain selectivity which leads to the preferential formation of the [C-pro α1(I)]$_2$[C-pro α2(I)] heterotrimer. These questions are addressed by energy minimization modeling of the individual C-propeptide structures, study of their docking interactions, and comparison of the heterotrimeric and homotrimeric C-pro structures and stability. The comparisons show the remarkable impact of the C-pro α2 chain on the structure of the assembled trimeric C-propeptide. In the modeling, the three chains were anchored and registered by a short C-terminal collagen triple-helical segment followed by the C-telopeptides in their docked conformation, and then the remaining C-propeptide chains were allowed to interact and dock. Surprisingly, propeptide trimerization did not proceed through the previously proposed N-terminal "oligomerization domain" of the C-propeptide [McAlinden et al. (2003) *J. Biol. Chem. 278*, 42200] but rather in the most C-terminal domains of type I procollagen chains. Molecular dynamics showed heterotrimer assembly to begin with dimer formation between globular G2α2 and the G2α1$_2$ domains followed by trimerization at the G1 domains. Assembly initiation in the putative oligomerization coiled-coil domain is not possible because of the Pro residues at positions 3, 7, and 11 at the N-terminus of the α2 C-propeptide chain. To confirm the computations and proposed assembly pathway, the G2α1 and G2α2 domains were prepared recombinantly as the maltose binding protein constructs, and their interactions were studied by dynamic light scattering and gel filtration chromatography. Under the conditions examined MBP remained as monomer, MBP-G2α1 and MBP-G2α2 alone formed dimers, but a 2:1 mixture of MBP-G2α1 and MBP-G2α2 favored trimer formation. Thus, the C-terminal globular domains (G2) of the type I collagen C-propeptides play a crucial role in the initiation of intermolecular assembly and heterotrimer selectivity.

The mechanism of the in vivo assembly of a three-chain procollagen molecule is one of the most interesting problems remaining with regard to collagen biosynthesis. This is particularly the case with the assembly of the heterotrimeric collagens, including the major collagen, type I. Each of the α chains of the heterotrimeric type I procollagen molecule has five major domains: the N-propeptide, N-telopeptide, triple helix, C-telopeptide, and C-propeptide. Each of these domains has a specific structure and specific function in one or more of the processes of molecular assembly, secretion, and the subsequent extracellular fibril assembly and stabilization. The proteolytically released extracellular N- and C-propeptides, as assembled trimeric molecules in their own right, have physiological functions independent of their direct effects as part of the procollagen molecule. Despite decades of work on type I collagen and the knowledge of the complete sequences of each α chain and of the chemistry of the cross-linkages, the structure of each domain is not known in terms of molecular conformation except in the most general terms.

The carboxyl propeptides of type I collagen pro α chains have been recognized as the key interactants in initiation of the formation of the trimeric procollagen I molecule in the C to N direction (*1*−*3*) within the rough endoplasmic reticulum (ER) (*4*−*6*). Correct interactions of the three C-pro α chains are crucial in producing the ultimately correct triple-helical collagens, as demonstrated by both lethal and nonlethal mutations within the C-propeptides that lead to defective collagens (*7*−*11*). There are three principal aspects of the heterotrimeric ([pro α1(I)]$_2$[pro α2(I)]) assembly problem: (1) the synthesis and elongation of the pro α1 and pro α2 chains within the same ER compartment; (2) the folding of the nascent chains and their trimerization; and (3) during trimerization, the selectivity of heterotrimer over homotrimer formation. The homotrimer ([pro α1(I)]$_3$) is found in small amounts in most type I collagen systems and is the default structure in cases where the synthesis of the α2 chain is blocked. Although, as shown by Kuznetsova et al. (*12*), the triple-helical [α1(I)]$_3$ part of the homotrimer is more stable than the corresponding heterotrimer triple-helical domain, the heterotrimer triple helix is the favored structure.

† This work has been supported by NIAMSD Grant AR-013921 (to A.V.).

* To whom correspondence should be addressed. Phone: 312-503-1355. Fax: 312-503-2544. E-mail: aveis@northwestern.edu.

‡ Present address: Department of Oral Biology, SUNY at Buffalo School of Dental Medicine, 320 Foster Hall, 3435 Main St., Buffalo, NY 14214.

10.1021/bi0508338 CCC: $30.25
Published on Web 10/29/2005

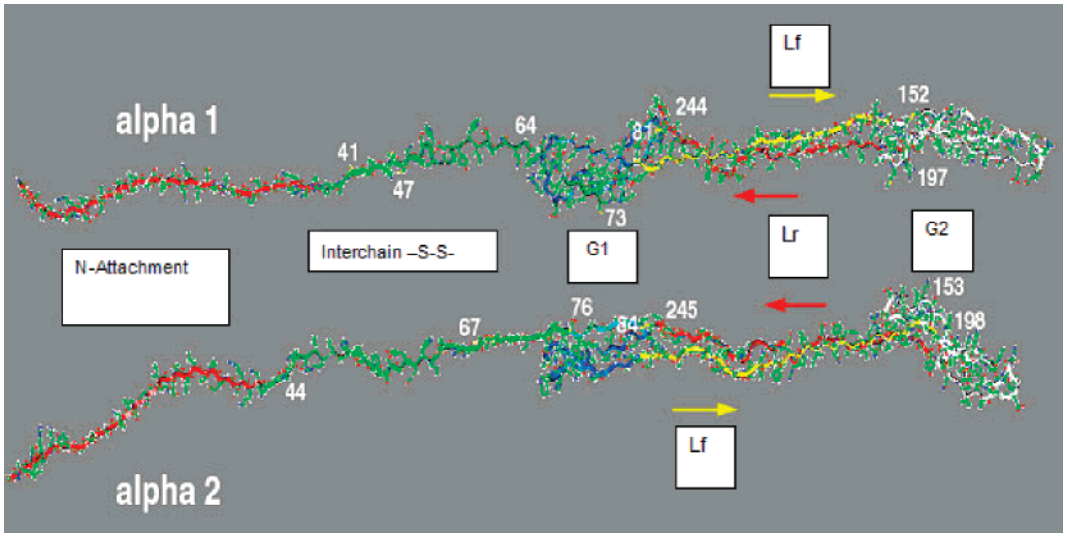

FIGURE 1: Preliminary modeling results of the separate C-pro α1(I) and C-pro α2(I) chains, N to C direction, defining the domains of different character (*13*). Each domain is labeled. As individual chains the N-attachment sequences and the interchain -S-S- sequences are depicted as in the near trans-extended form, but they are of low stability and will behave in solution as random chain polypeptides. They were found not to favor the α-helix conformation. The Lf and Lr segments likewise have only partial *β*-strand structures, as pointed out in ref *13* and are more likely to behave as less extended chains in solution. Thus, these models represent a schematic view of the real chain structures.

The structures of the C-propeptide domains and their roles in chain selection and folding are not known. These may be particularly important in molecular assembly since less homotrimer is secreted into the extracellular space.

Our initial modeling study had determined that the highly homologous, but distinctly individual, human C-pro α1(I) and C-pro α2(I) chains would contain several domains (Figure 1) (*13*). The labeling of the domains is indicated in the figure and is color coded. The N-telopeptide attachment domain (in red) does not have any defined structure in either chain but is depicted as virtually trans-extended. The next domain (in green) is also an extended chain but contains the potential interchain disulfide bond interaction zone. The extended chain is followed by a complex folded region that is comprised of three segments. The chain folds into the first globular segment (G1, in blue) and then exits to an extended chain-linking region that contains *β*-structure ($Lf_{forward}$, in yellow). The chain folds to a second globular segment (G2, in white) and then reaches back through a final extended chain *β*-structure ($Lr_{reverse}$, in red) linking G2 to G1. The folding of G2 is stabilized by an intrachain disulfide bond, α1 152−197 or α2 153−198, while the entire folded region is stabilized by another intrachain disulfide bond, α1 81−244 or α2 84−245. The modeling computations showed that the Lf-Lr linking region had sequences in antiparallel *β*-sheet conformation. All of the Cys residues are numbered in Figure 1. The models shown were developed utilizing the known pattern of intramolecular disulfide bonds (*14*) in the C-propeptides and the observation (*15*) that in globular proteins there is a very close relationship between disulfide-bond patterns, protein structure, and sequence homology despite specific sequence differences. Overall, the C-pro α1(I) and C-pro α2(I) show 61% identity and 21% conservative homology.

Neither G1 nor G2 have any major sequences forming α-helical segments but contain *β*-turns. The structures in Figure 1, computed independently of each other by energy minimization procedures, are very unlike the C-Pro domains of type IV (*16*) and XVIII (*17*) collagens whose structures were determined by direct crystallographic analysis. No comparable studies of the fibrillar collagen C-propeptides are available.

Development of the main triple-helical collagen fold domain requires only the alignment or registration of the appropriate three $(GXY)_n$ chains. Full-length or shortened sequences will suffice in vitro or in cells as long as an alignment domain is provided (*18*, *19*). Even α2(I) chains will form homotrimeric triple helices if suitably anchored and registered. In the normal in vivo situation, this anchoring and registration role is the function of the C-propeptides. Myllyharju et al. (*20*) have convincingly demonstrated that in cells C-pro α1(I) can form homotrimers or heterotrimers with C-pro α2(I) chains but that the C-pro α2(I) chains cannot form homotrimers. These data suggest that the information leading to the formation of the heterotrimer is a property intrinsic to the C-pro α2(I) propeptide domain, but the features responsible are not known. Resolution of this problem is not straightforward. Smith and colleagues (*21*, *22*) have shown that a mutation removing the 10 C-terminal amino acids of C-pro α2(I) (*21*) or the removal of the G2 intrachain disulfide bond (corresponding to human α2 153−198) from the C-pro α2(I) (*22*) disrupted assembly with C-pro α1(I). However, we have shown in yeast two-hybrid studies that the intact C-pro α2(I) propeptides are capable of interacting with each other (*13*) to form dimers. Further, in vitro interaction studies with recombinant C-pro α1(I) and C-pro α2(I) constructs suggested that the propeptides reacted differentially with ER resident chaperone proteins.

Thus, two important questions emerged: how do these two homologous structures interact to form the three-chain complete propeptide and how does the system select for either the heterotrimer or homotrimer? Alternate formulations of these questions are how do the three chains dock together and what is the conformation of the C-pro α2(I) propeptide as compared to that of the C-pro α1(I) propeptide? Studies thus far in other laboratories have been directed to analyses

of the known sequences and prediction of the structures in terms of typical protein conformations, such as extended β-sheets or α-helical segments (*23*, *24*), and how these form oligomerization zones (*25*, *26*). The three-chain coiled-coil oligomerization zones proposed for collagen were in the N-terminal sequences of the C-propeptides, joining to the C-telopeptide. The three coiled coils were presumed to be in the α-helical conformation. In the models depicted in Figure 1 that domain of the α1 C-propeptide did not show any tendency to form an α-helix as an independent chain, and the α2 C-propeptide chain contains several Pro residues which would prevent α-helix formation in the N-terminal domain (*27*). Thus, the type I collagen does not seem to fit into the pattern proposed for the other homotrimer fibrillar collagens. Small-angle X-ray scattering studies have examined the type III homotrimer C-propeptide (*28*). These SAXS studies have shown it to have a trilobed structure emanating from the C-telopeptide, but the Kratky-based SAXS models were at low resolution and did not provide the detailed conformational information needed to determine the nature of the controlling interactions.

In the present work we have not imposed any preconceived conformation ideas but have updated the energy minimization modeling procedures used earlier for each strand (*13*), added information on the placement of the disulfide bonds, and docked the three chains together. We have carried this out for both the [C-pro α1(I)]$_2$[C-pro α2(I)] heterotrimer and the [C-pro α1(I)]$_3$ homotrimer. The models show the remarkable impact of the C-pro α2 chain on the structure of the assembled trimeric C-propeptide and suggest that chain selection and trimerization may not begin at the N-terminal "oligomerization domain" of the C-propeptide. We have carried out interaction studies using recombinant constructs of the putative selection and interaction domains by dynamic light scattering and gel filtration chromatography to confirm the modeling-based predictions.

MATERIALS AND METHODS

Modeling Protocols. As in the previous modeling studies of the propeptides (*13*) and telopeptides (*29*, *30*), a suite of programs for protein structure analysis from ACCELRYS/MSI, including BIOPOLYMER, DISCOVER, and INSIGHT II, was used for all computations with a SGI Octane R12000 SE computer. The sequences used were all obtained for human type I procollagen from the latest updates in the SWISS PROT Data Bank, Accession Number P02452 for α1(I) and Accession Number P08123 for α2(I). Once the appropriate sequences were loaded, the consistent valence force field (CVFF) algorithms (*31*) were used to minimize the total energy. Reiterations of the derived structure were continued until the root-mean-square deviation (RMSD) was below 0.0001. All of the parameters of the energy-minimized structures will be made available as .PDB files upon request. After a structure was determined, then that structure was relaxed at 300 K for a few hundred picoseconds, and molecular dynamics was used to confirm that the energy-minimized structure was not in a false or intermediate energy conformation. The effects of point mutations in sequence could be determined with relative ease by taking the parameters determined for a particular "domain" sequence and then substituting a mutant amino acid and recomputing the energy. This did not require extended reiterations in most cases.

Plasmid Construction. Sequence information from human pro α1(I) and pro α2(I) plasmids, from pHHα1 and pHHα2 (*13*), respectively, was used to design primers to amplify the G2 regions of C-pro α1(I) and C-pro α2(I). The forward and reverse primers included a *Bam*HI and *Sal*I site, respectively. The conditions for the PCR amplification were 94 °C for 3 min, followed by 30 cycles of 94 °C for 30 s, 60 °C for 1 min, and 72 °C for 1 min. This was followed by an extension of 72 °C for 10 min. Following amplification, the amplified products were digested with *Bam*HI and *Sal*I and cloned in the *Bam*HI, *Sal*I site of the maltose binding protein expression vector pMal-p2E using standard conditions (*32*) to generate the plasmids pMal-p2E-G2α1 and pMal-p2E-G2α2. The signal sequence of the *malE* gene on this vector is intact, so fusion proteins are exported to the periplasm, allowing them to fold correctly (*33*). The pMAL-p2 vectors are an especially good choice if it is known that the protein of interest is a secreted protein.

Expression and Purification of G2 Domains of C-pro α1 and C-pro α2. The recombinant plasmids were cloned into bacteria BL21(DE3). A single colony from each construct was selected and inoculated into 50 mL of Luria−Bertani broth (LB) containing 50 μg/mL ampicillin and grown overnight at 30 °C. The next day, the culture was transferred to 500 mL of LB containing 0.2% glucose and 50 μg/mL ampicillin and grown at 30 °C until the OD_{600} was 0.6 (about 4 h), and isopropyl thiogalactoside (IPTG) was then added to the remainder to a final concentration of 0.5 mM, and incubation was carried on for 16 h.

For the purification of the fusion proteins at the end of the 16 h induction, the cells were recovered by centrifugation at 5000*g* for 30 min. The cells were washed once with ice-cold 20 mM Tris-HCl, pH 7.4, and resuspended in 100 mL of 20 mM Tris-HCl, pH 7.4, containing 200 mM NaCl and 0.5 mM EDTA (column buffer). The suspension was freeze-thawed three times to lyse the cells, sonicated to shear the DNA, and then centrifuged at 12000*g* for 15 min at 4 °C. The supernatant was then passed over an amylose−Sepharose column, equilibrated with column buffer, washed with 25 bed volumes of column buffer, and then eluted with column buffer containing 50 mM maltose. The eluted protein was dialyzed against 20 mM Tris-HCl, pH 7.4, and loaded on a Bio-Gel TSK-DEAE-5-PW column. After the protein was bound, the DEAE column was washed with the 20 mM Tris-HCl, pH 7.4, loading buffer and then eluted at a flow rate of 5 mL/min using a Beckman System Gold HPLC system with a 0−1 M NaCl salt gradient. The protein peak was pooled, dialyzed against 20 mM Tris-HCl, pH 7.4, and then concentrated to 1 mg/mL using a Millipore Centriplus YM-10 concentrator. These stock solutions (MBP-G2α1, MBP-G2α2) of 1 mg/mL in 20 mM Tris-HCl at pH 7.4 were used to prepare all dilutions for the interaction studies. SDS−gel electrophoretic analysis of the final solutions was carried out in the presence and absence of β-mercaptoethanol.

Gel Filtration of G2α1, G2α2, and Mixtures of G2α1 and G2α2. The stock solutions were diluted to 0.5 mg/mL in 20 mM Tris-HCl and 0.15 M NaCl at pH 7.4. An aliquot of each solution was loaded individually onto a Bio-Sil SEC 250 molecular sieve chromatography column. The column was run at a flow rate of 1 mL/min in the same solvent

composition. Fractions of 1 mL were collected and correlated in time to the chromatogram. The peak corresponding to the monomer was collected and concentrated, and the concentration was estimated. These G2α1 and G2α2 monomer peaks were then concentrated with the Centriplus YM-10 concentrator and made up to new 1 mg/mL stock solutions in 20 mM Tris-HCl and 0.15 M NaCl at pH 7.4. An aliquot of each monomer solution was chromatogramed again. A mixture at a 2:1 (MBP-G2α1:MBP-G2α2) weight ratio in the same solvent was prepared, at final concentrations of 0.5 and 0.25 mg/mL, respectively. After 30 min equilibration, 100 μL of the mixture was loaded on the molecular sieve column and run as before.

Dynamic Light Scattering Analysis. MBP-G2α1 and MBP-G2α2 in 20 mM Tris-HCl and 0.15 M NaCl, pH 7.4, stock solutions, as described above, were used for evaluation of the distribution of diffusion coefficients, *D*, at 20 °C. Although the ultimate intent was to examine mixtures in the 2:1 weight ratio, MBP-G2α1 and MBP-G2α2 were both examined independently at 1.0, 0.75, 0.5, and 0.25 mg/mL. Over this range of concentrations the weight fraction distribution of *D* for each protein was essentially constant. Thus, the mixing experiments were carried out at concentrations of 0.5 mg/mL MBP-G2α1 plus 0.25 mg/mL MBP-G2α2 in 500 μL of 20 mM Tris-HCl and 0.2 M NaCl at pH 7.4. The G2α1 or G2α2 solutions or the mixtures were examined in 5 mm^2 Starna cells and analyzed by dynamic light scattering at 20 °C using a temperature-controlled PD 2000 DLS (Precision Detectors, Franklin, MA). The laser wavelength was 800 nm, and the scan parameters were as follows: smoothness 20; sample time 15 ms; and run time 5 s with 50 accumulations. Each sample was analyzed 10 times, and the average of the percentage distribution across the diffusion coefficients was calculated.

RESULTS

Basic Assumptions for Modeling and Docking. In the initial study of the separate α1 and α2 C-propeptides (*13*), the sequences had been started at the N-terminal residue at the C-proteinase cleavage sites in each chain. The straightforward energy minimization process yielded the two different, but overall very similar domain structures shown in Figure 1. In the present computations, where alignment, docking, and folding of the three-chain structure was the objective, each chain sequence began with a nine amino acid sequence capable of participating in an in register triple-helix segment characteristic of the collagen domain. The correct sequence for the C-telopeptide and C-propeptide of each chain was then grafted onto the $(Gly\text{-}Pro\text{-}Hyp)_3$ sequence, producing a polypeptide chain in the order -$(Gly\text{-}Pro\text{-}Hyp)_3$-C-telopeptide-C-propeptide. The alignment and assembly of the three chains were initiated by building the collagen triple-helix portion in the α1-α2-α1 chain stagger (*29*). The $(Pro\text{-}Pro\text{-}Gly)_n$ triple-helical parameters for this model were downloaded from high-resolution (1.4 Å) X-ray crystallographic coordinates of the Protein Data Bank, ID 1k6f (*34*). The remainder of each chain (TELO + PRO) was then joined to the C-terminus of the triple helix in the proper stagger to register the chains. Although the putative structure of the assembled three-chain telopeptide domain had been determined (*30*) for the free collagen molecule (sans C-propeptide), it was evident in that work that the structure of the telopeptide domain might be different when the C-propeptides were attached. Nevertheless, to simplify the computations for these formidably large data sets, the α1 and α2 C-propeptides, with 246 and 247 amino acids, respectively, were entered in the conformations determined earlier (*13*), as presented in Figure 1, grafted onto the C-telopeptides in their three-chain assembled conformations (*30*). The intrachain disulfide bonds of the C-propeptides were in place, but no other interchain interactions were specified. The structure of the composite, chain registered structure was then determined by an extended minimization procedure. Interchain interactions ultimately led to a specific docked structure that allowed the formation of interchain disulfide bonds as in the native intact type I collagen C-propeptides. Interchain disulfide bonds, as specified in the literature (*14*, *35*, *36*), were then introduced, and the conformations were determined again. The additional constraints of the interchain disulfide bonds led to relatively minor changes in the chain conformations. As anticipated, the final minimum energy conformation changed in all domains in the assembled C-telopeptide and C-propeptide. Although the N to C registration direction is not the physiologically correct assembly pathway, this approach has proven to be an informative way to see the relative roles of the different domains in achieving their final conformation in the assembled three-chain structure. As shown below, it became evident that the trajectory of folding had a number of false energy minima. These intermediate states were followed through molecular dynamics.

Both C-propeptides have a single N-linked glycosylation site, at N148 in α1 and N149 in α2 (*37*−*39*), just at the end of the Lf domain four residues preceding the S153···S198 disulfide bond that stabilizes the G_2 domain (13, Figure 1). The carbohydrate units consist of a branched chain of mannose residues linked through two residues of *N*-acetylglucosamine to the N148 or N149 glycosylation site. The number of mannose residues is variable in the C-propeptide; in chick the C-pro α1 was found to contain 9 mannose residues while the C-pro α2 contained 13 (*40*). Since the glycosylation site is in the linking region between G1 and G2 rather than either globular folded region, we have, as a first approximation to the structure, not included the glycan units in the computations. This was considered to be reasonable on the basis of the finding by Lamande and Bateman (*39*) that unglycosylated mutant pro α1(I) folded and assembled into trimeric molecules with pro α2(I). However, in that study the C-pro α2(I) glycosylation was normal. Some subtle differences were noted in the subsequent proteolytic processing of the secreted protein. The problem will be reconsidered in the Discussion.

Modeling Results. After registration of the three propeptide units by forming the N-terminal triple helix, but with no other interchain constraints, as schematically illustrated in Figure 2, left panel, the interaction between chains was monitored as the compound three-chain assembly structure was developed. In the cell, the docking is presumed to occur within the ER compartment, with assembly in the C to N direction after the individual chains are completely elongated and released from their ribosomal synthesis. It is also presumed that each chain has folded its G1 and G2 domains with the intrachain disulfide bonds established. Thus the N to C interaction protocol used here is not comparable to that in a cell. Nevertheless, three interesting stages of assembly were

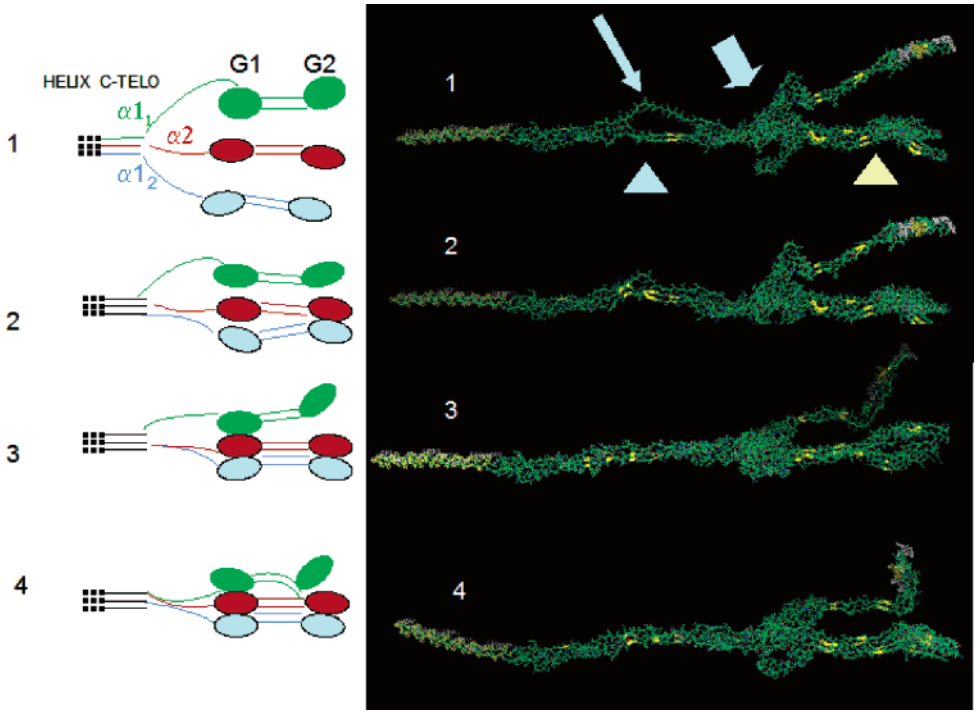

FIGURE 2: Stages during the assembly of the heterotrimer C-propeptide, showing the development of folding and domain interaction. Chains were anchored and registered at the C-terminal triple helix domain in the $\alpha1_1$-$\alpha2$-$\alpha1_2$ order. Left, schematic view; right, modeling results. Stages 1 and 2: The $\alpha1_2$ and $\alpha2$ chains begin to interact almost immediately at the G2 and G1 domains; folding and interaction proceed more slowly through the interchain disulfide bonding and N-attachment domains. Stage 3: The G1 trimer is formed, but G2$\alpha1_1$ does not participate. Stage 4: At the minimum energy conformation, some interaction of $\alpha1_1$G2 with the G2$\alpha2$-G2$\alpha1_2$ becomes apparent, but the $\alpha1_1$Lrf remains independent. The folding becomes more ordered throughout the N-attachment and interchain disulfide bonding domains. Key: long arrow, the putative trimerization zone (*26*) immediately C-terminal to the junction with the C-telopeptides; short thick arrow, aggregation to trimer in the G1 domains; blue arrowhead, alignment and interaction of $\alpha2$ and $\alpha1_2$ chains in the sequence immediately C-terminal to the junction with the C-telopeptide; cream arrowhead, interaction between G2$\alpha2$ and G2$\alpha1_2$ and their Lfr domains.

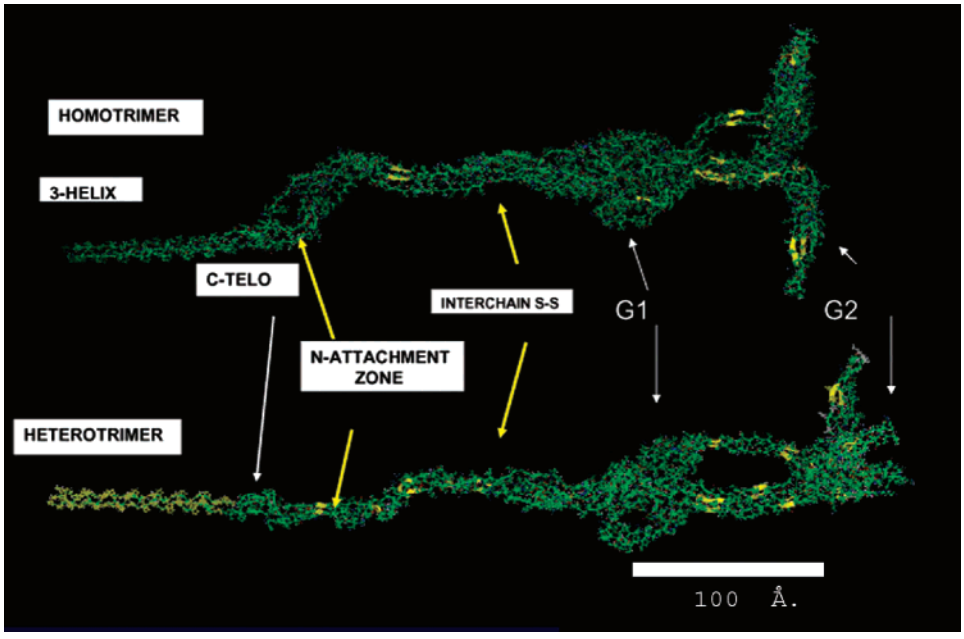

FIGURE 3: Comparison of the final energy-minimized structures of the homotrimer (top) and heterotrimer (bottom). The differences in the N-attachment and G2 domains are evident. In contrast to structural changes noted with molecular dynamics at various stages in the energy minimization process, these final structures did not change after a brief period of relaxation at 300 K and refolding.

noted. These are illustrated in the series of developing models shown in Figure 2, right panel. In the earliest assembly stage, 1, the docked C-telopeptides had already changed conformation as a result of bearing the attached C-propeptides, but it was apparently difficult to drive the three-chain interaction through the N-terminal part of the C-propeptide (long arrow). However, it is clear that, in the $\alpha1_1$-$\alpha2$-$\alpha1_2$ chain stagger arrangement, the $\alpha2$ and $\alpha1_2$ chains in that segment of the C-propeptide do interact readily (arrowhead, blue). Trimerization appears to initiate by interaction of all three of the G1 domains (short arrow), but the two-chain interaction between $\alpha2$ and $\alpha1_2$ is resumed in the Lfr and G2 domains

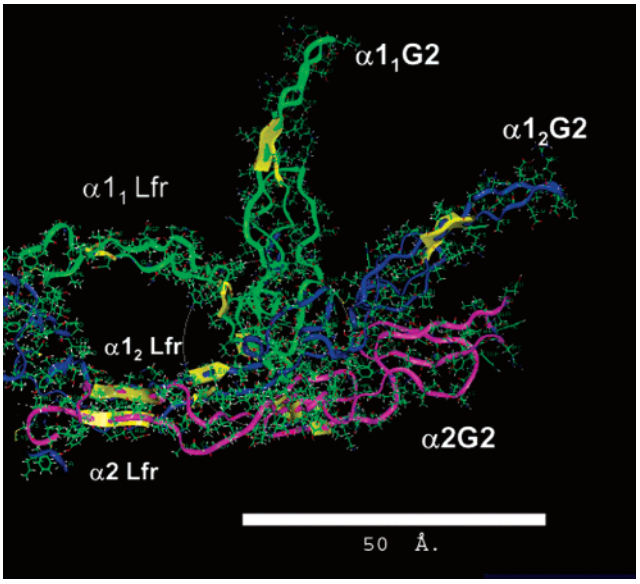

FIGURE 4: Detailed view of the assembled G2 domain of the heterotrimer as located at the end of the assembled C-propeptide. As indicated, $\alpha1_1$ is in green, $\alpha1_2$ is in blue, and $\alpha2$ is in magenta. The Lfr regions and G2 regions of $\alpha2$ and $\alpha1_2$ are tightly interacted, while the Lfr and G2 of $\alpha1_1$ do not interact except near the $\alpha1_2$ and $\alpha2$ Lfr-G2 junction.

(arrowhead, cream). It is likely that, in the real situation within the ER compartment, C ⟹ N dimerization of the $\alpha2$ and $\alpha1_2$ nascent chains begins in G2 as the first stage of interaction. Formation of the more C-terminal (G2$\alpha2$-G2$\alpha1_2$) dimer may thus be at the root of chain selection for heterotrimer formation. Successive stages of relaxation and structure perfection reduce the energy of the overall structure (Figure 2, panels 2, 3, and 4) but do not produce a coiled-coil structure within the putative "trimerization domain" (*26*). Trimerization appears to be established between the ($\alpha2$-$\alpha1_2$) G2-Lfr dimer and the free $\alpha1_1$ via the G1 domain interactions. The final heterotrimer structure achieved when the minimization was to a RMSD of 0.0001 is shown at the bottom in Figure 3. A closer look at the G2 region of the heterotrimer (Figure 4), turned and color coded so as to more nearly see an axial view, shows the marked difference in the spatial disposition and interactions of G2$\alpha1_1$ and G2$\alpha1_2$ with G2$\alpha2$ and the dimeric nature of the $\alpha1_2$ Lfr -$\alpha2$ Lfr connecting domain, reaching to G1. Figure 5 emphasizes this relationship as well as to the G1 domains.

The energy-minimized models shown here include the telopeptide and a region of triple helix, appearing to make the whole structure extended and rodlike. However, the dimension from G2 to the foot of G1 and folded -S-S- regions is 107.2 Å. The longest dimension computed from the low-resolution X-ray diffraction data on the type III C-propeptide (triple helix and C-telopeptide excluded) was on the order of 110 Å (*28*), entirely compatible with the folded model presented here. In the free propeptide in solution, where the N-terminal domain is not attached, it is likely that the joining sequence would be flexible.

Identical computations were carried out for the construction of the [C-pro $\alpha1(I)_3$] homotrimer with all of the same assumptions (Figure 3, top). The final energy of the homotrimer was higher (less stable, +19000 kcal/mol) than that of the heterotrimer (−1200 kcal/mol), and the structure, particularly in the region joining to the C-telopeptide, was much less compact (Figure 6). Compared with the same domain of the heterotrimer, the G2$\alpha1$ domains of the homotrimer did not come together as in the heterotrimer (Figure 7).

We have not considered the role of the glycosylation of the C-propeptide chains since, as noted earlier, Lamande and Bateman (*39*) had shown that nonglycosylated C-pro $\alpha1$

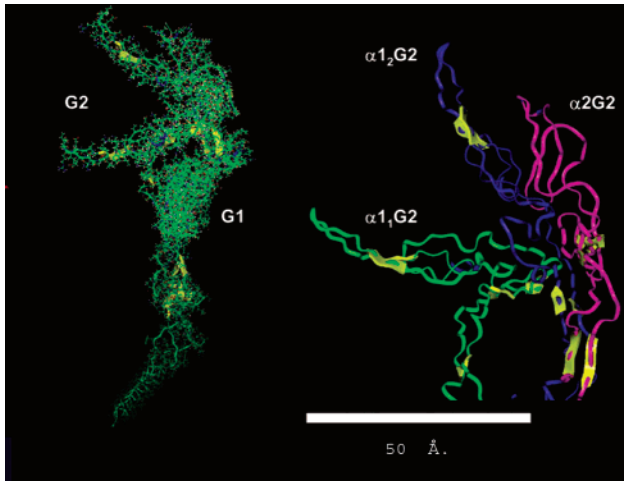

FIGURE 5: View of the entire heterotrimer from the end of the triple helix to the G2 domains, from the perspective of looking obliquely down from the G2 domain (left panel). The G2 domain backbone structure is shown on the right. The G2-Lfr dimerization of the $\alpha2$-$\alpha1_2$ chains is evident. The scale bar refers only to the right panel.

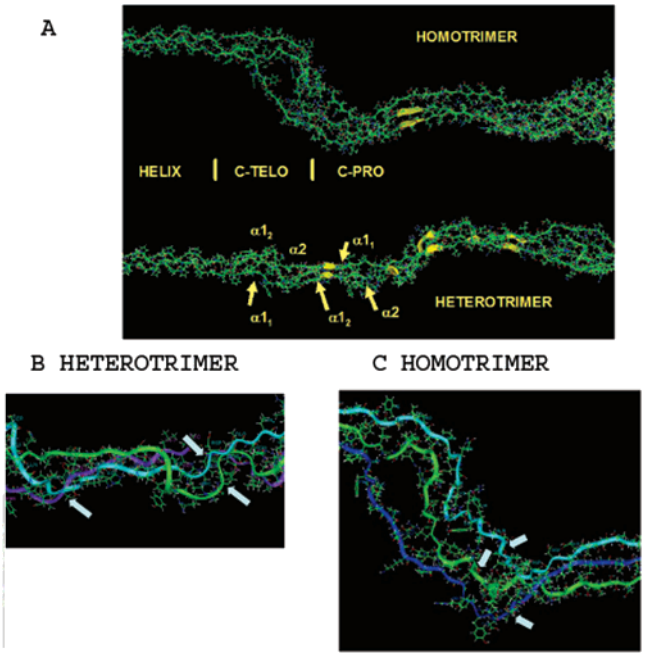

FIGURE 6: (A) Comparison of the C-telo-C-pro connecting regions of the homotrimer (top) and heterotrimer (bottom). In the heterotrimer the $\alpha1_1$, $\alpha1_2$, and $\alpha2$ chains are indicated by the thin arrows. The thick arrows along the chains denote β-structure. Note that the homotrimer is much less well ordered in this region with only a single region of parallel β-strand between two of the three identical (but staggered) chains. (B) Magnified view of the junction between telopeptide and propeptide sequences of the heterotrimer. The arrows show the cleavage points of the C-proteinase. (C) Magnified view of the junction between telopeptide and propeptide in the homotrimer. The arrows show the C-proteinase cleavage points. The change in folding because of the presence of the $\alpha2$ chain is evident. See also Figure 12.

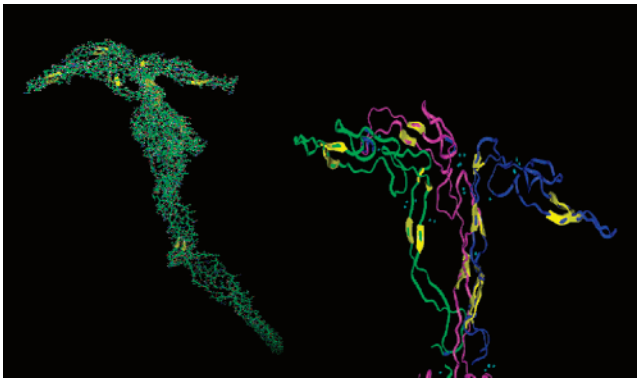

FIGURE 7: View of the homotrimer G2 domain comparable to that in Figure 5. The difference in orientations of the G2 domains is striking, indicative of their minimal interaction.

chains assembled into heterotrimers with C-pro α2. However, the glycosylation site in all three chains is in Lfr just at the junction between G2 and Lfr domains. The modeling shows clearly (Figures 4 and 7) that all three Lfr strands do not join in either the homotrimer or heterotrimer, leaving space for addition of the bulky complex carbohydrate moieties.

Interaction Studies: Experimental Tests of the Predictions. There are several ways to test these models. The most direct way is to make the recombinant propeptides, allow the constructs to fold, and then study the resulting structures by X-ray crystallographic means. That work is in progress and some crystals have been obtained, but no structures have been determined as yet. An alternative was to make the recombinant domains and then study their interactions. The first interaction study (*13*) examined the full-length C-propeptides in the yeast two-hybrid system. Those experiments showed unequivocally that C-pro α1 and C-pro α2 could interact with each other and with themselves in the yeast cells to allow expression of the *β*-galactosidase marker. Further studies with the same constructs prepared in a rabbit reticulocyte lysate system and examined in a variety of affinity chromatography systems also showed strong interactions among the propeptides. Deletion peptide constructs were then prepared, eliminating either the first 37 residues, now denoted in Figure 1 as the N-attachment domain (ΔN), or the last 36 residues comprising the C-terminal Lr chain (ΔC), joining G2 to G1. The α1-ΔN and α1-ΔC constructs retained the ability to bind to intact C-pro α1 and C-proα2. The α2-ΔN behaved similarly, but α2-ΔC lost its ability to bind to both intact C-pro α1 and C-proα2. These data were supported by concurrent deletion studies by Doyle and Smith (*22*). The modeling studies reported above suggest that either the C-pro α2 Lr or G2 domains appear to be crucial for the selection and initial dimer formation process, whereas the trimer formation involved the G1 domain interactions. Therefore, the constructs representing the α1-G2 and α2-G2 domains were prepared and their interactions studied.

Dynamic Light Scattering. Maltose binding protein (MBP) constructs were selected for the interaction study because the MBP, by itself, is very soluble and is present in solution as a monomer. In each case, the constructs recovered from the maltose binding column were collected, concentrated, and run over a DEAE column from which they eluted as the major peak. As shown in Figure 8, gel electrophoresis of each construct recovered from the DEAE chromatography in a nonreducing but denaturing buffer showed that neither

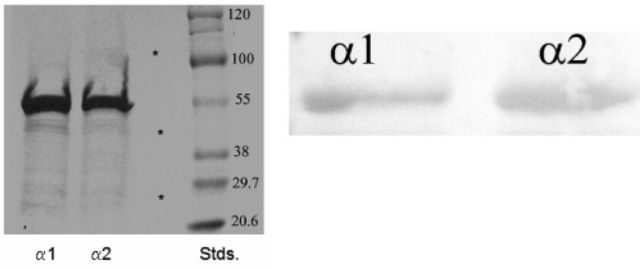

FIGURE 8: Gel electrophoresis of the MBP-G2α1 and MBP-G2α2 maltose binding protein constructs after collection on the DEAE column. Left panel: Gel run under denaturing and nonreducing conditions. The asterisks denote the presence of a low level of some larger and smaller components in the isolate. Right panel: Western blot of the gel using a mixture of antibodies to C-pro α1 and C-pro α2. There is little cross-reactivity between the antibodies.

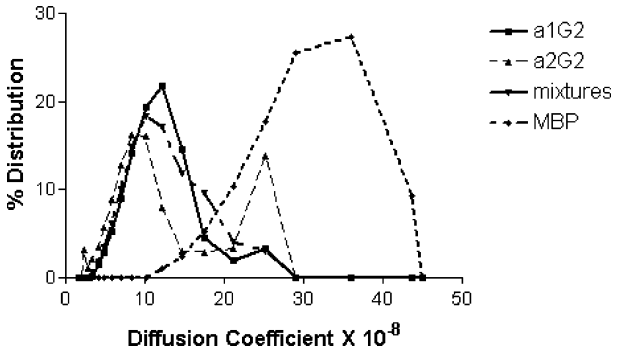

FIGURE 9: Distribution of diffusion coefficients, D^{20} (cm^2 s^{-1}), of the components of solutions of the MBP constructs (∼52000 Da). The MBP-G2α2 has the highest content of monomer. This is almost entirely brought into a trimer when mixed with MBP-G2α1 at a 1:2 ratio. The MBP, with a lower molecular mass (40200 Da) is mainly monomer. These data clearly show the tendency of the MBP constructs to interact. When mixed with MBP-G2α1, the monomer present in the MBP-G2α2 is greatly diminished, showing a preferential interaction between MBP-G2α1 and MBP-G2α2.

the MBP-G2α1 nor the MBP-G2α2 construct formed intermolecular disulfide bonds, and the principal components had apparent molecular weights of ∼52000. The formula weight of the MBP-α1G2 is 51860. Small amounts of both larger and smaller proteins were present (Figure 8, *). When the DLS measurements were carried out on the DEAE-recovered MBP-G2α1 and MBP-G2α2 in pH 7.4, 0.15 M NaCl solutions, however, it was obvious that aggregates were present. A detailed study of the peptide interactions at higher and lower salt concentrations, pH, temperature, and protein concentration was carried out but will be reported elsewhere. Here we have chosen to present these data at near physiological ionic strength. The distributions of weight percent protein as a function of D are shown in Figure 9 for the MBP-G2α1 at 0.5 mg/mL and MBP-G2α2 at 0.25 mg/mL and for the mixture at the same concentrations (a molar ratio of α1:α2 = 2:1). The values of D were neither corrected for solution viscosity nor extrapolated to zero concentration, but since all measurements were at the same concentration, buffer composition, and temperature, the shift in D distributions to lower values represents the aggregation of the molecules in solution. Monomeric MBP alone (MW = 40500) is an asymmetric molecule with a radius of gyration of 22.0 Å (*41*) and dimensions of 30 × 40 × 65 Å (*42*). In our study, recombinant MBP (1 mg/mL), purified as above, has $D^{20} = 3.3 \times 10^{-7}$ cm^2/s. All of the distributions of D

for the G2 constructs yielded biphasic distributions with a small "monomer" peak at $D^{20} = 2.5 \times 10^{-7}$ cm^2/s and larger amounts of aggregate. The differences noted were reproducible and showed that although the MBP by itself did not dimerize, both MBP-G2α1 and MBP-G2α2 formed aggregates to different extents. Moreover, at 20 °C, solutions tended to age and the aggregate level increased. When the proteins were mixed in the 2:1 molar ratio, almost all monomer in the solution was depleted and a broad distribution of lower *D*, indicating aggregates, was observed. To follow this further, we resorted to gel filtration analysis of the molecular species.

The G2 adducts are linked to the C-terminal residue of the MBP, which is along the outer surface of the molecule and undoubtedly increases the molecular asymmetry, increasing the frictional coefficient and decreasing *D*, but this accounts for only a part of the observed decrease in *D*. If one assumes that the dimers or trimers were formed, but the overall shape did not change, then *D* values would be expected to vary according to $M^{-1/3}$. Thus if $D_{monomer}$ is 2.5×10^{-7} cm^2/s, then D_{trimer} should be reduced to 1.7×10^{-7} cm^2/s, so the nonspecific or misfolded aggregates making up the bulk of the solution must be larger than trimer, or very asymmetric. Further studies would be required to explore that behavior.

Molecular Sieve Chromatography. To circumvent the time- and solvent-dependent aggregation problem, the initial stock solutions of the MBP constructs obtained from the DEAE columns were passed over a Bio-Sil SEC 250 molecular sieve column that had been calibrated with molecular weight standards, and the peak fractions were collected. As anticipated from the DLS data, the monomer content was quite low compared to the large aggregates [Figure 10, A (α1), B (α2)]. However, the presumed monomer peaks were collected, concentrated as described above, and then the chromatography was repeated immediately on the same column under the same running conditions. The collection, concentration, and rechromatography were accomplished in about 2 h. These data showed that, in this brief time span, the monomer peaks remained as monomer (10.2 min elution), but dimer (9.2 min) and trimer (8.3 min) peaks were seen [Figure 10, C (α1), D (α2)]. Thus, while it appears that the interaction represents a dynamic situation, the immediate rechromatography of an individual peak still provided an adequate representation of the interaction state and fraction composition. Most strikingly, however, when the monomer peaks were mixed in a 2:1 (α1:α2) ratio and then chromatographed, there was a shift in the elution profile showing an enhanced trimer content. Compare panels C and D of Figure 10 showing monomers of α1 and α2 rechromatographed individually to Figure 10F where they were mixed in a 2:1 ratio before chromatography. This is emphasized in Figure 11, which compares the areas under the various peaks in the 2:1 mixture, with the calculated values that would be obtained if there were no interaction or aggregation. It is evident that the major components in the mixture were the trimer and higher aggregates and that, at the pH, ionic strength, and concentrations examined, little new dimer was formed. Recombinant MBP without G2 domains remained as an essentially lower mass monomer (10.6 min) (Figure 10E).

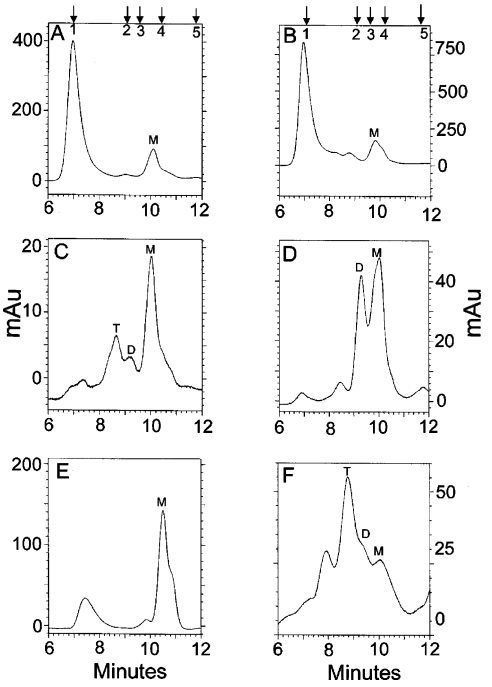

FIGURE 10: Molecular sieve chromatography of the MBP-G2α1 and MBP-G2α2 preparations. (A) Initial chromatography of MBP-G2α1 as recovered and desalted in the DEAE chromatography. (B) Initial chromatography of MBP-G2α2 as recovered and desalted in the DEAE chromatography. (C) Rechromatography of the peak collected from 9.5 to 10.5 min in (A), showing the monomer, dimer, and trimer content. (D) Rechromatography of the peak collected from 9.5 to 10.5 min in (B), showing the monomer, dimer, and trimer content. (E) Chromatography of recombinant MBP alone. (F) Chromatography of a 2:1 mixture of the contents of the 9.5−10.5 min peaks of (A) and (B).

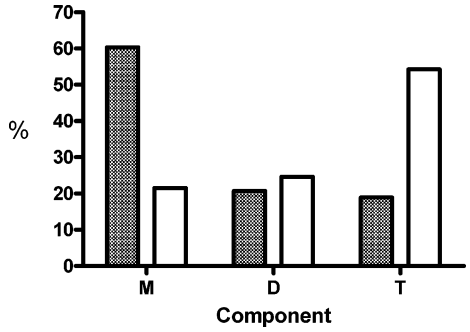

FIGURE 11: Relative contents of monomer, dimer, and trimer components in the MBP-G2α1, MBP-G2α2, and 2:1 mixture. The open bars represent the actual relative concentrations of these components derived from the chromatography in Figure 10F, whereas the hatched bars represent the components theoretically present if the components in panels C and D of Figure 10 did not interact. Obviously, the trimer content increased over that predicted in the mixture for the case of no interaction.

DISCUSSION

One of the most fascinating questions about collagen is why the most prevalent collagen of the body is the type I heterotrimer when it is now clear that the major (Gly-Xaa-Yaa)$_n$ domain of both the α1(I) and α2(I) chains can form stable triple helix homotrimers as long as the registration function can be performed (*19*). This is still more interesting from the perspective of the evolution of fibrillar collagens, since studies have suggested that the α2(I) chain was the early evolutionary precursor of the α1(I) chain (*43*) and the

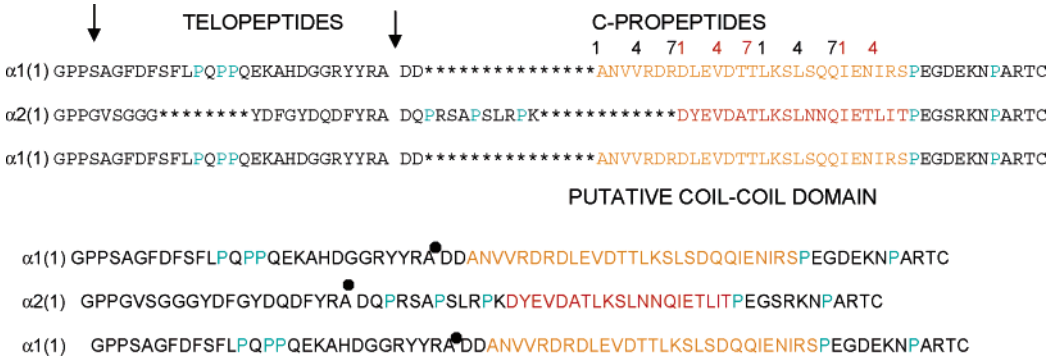

FIGURE 12: Sequences in the region connecting the telopeptides and C-propeptides in the α1 and α2 chains. Top panel: Sequence alignment if the alignment is based on the Col-Telo junction, the C-proteinase cleavage sequence, and the heptad repeat trimerization zones. The putative heptad repeats (*26*) are numbered for the α1 chain. Bottom panel: Sequence alignment based on the structures (see Figures 3 and 5) with the proper alignment of the C-terminal region of the triple helix. Note how the insertions of Pro residues in the telopeptide of α1 and the initial part of the heptad repeat segment of C-pro α2 result in the difficulty of organizing these segments in the heterotrimer and homotrimer. The proteinase cleavage points between telopeptide and propeptide are denoted by ●. The consequences of the marked shortening of the α2 C-telopeptide in terms of both conformation and cleavage by the C-proteinase are graphically illustrated in Figure 6.

fact that the triple-helical portion of the α1(I) homotrimer is more stable and has a higher melting temperature (~2.5 °C) and denatures more slowly at a rate about 100 times slower than the same region of the heterotrimer (*12*). The equilibrium unfolding temperature for the native heterotrimer is several degrees below the ambient body temperature (*44*), and this must be important physiologically (*45*). Thus, the α2(I) chain seems to be crucial for the proper in vivo function of the type I collagen. We have suggested (*27*, *29*) that the participation of the α2(I) sequences in the C- and N-telopeptides also provides important intermolecular interaction and cross-linking information for fibril organization. In the present work the focus has been on the role of the C-pro α2(I) in the selection and registration of the heterotrimer.

The modeling of the docked or assembled hetero- and homotrimers confirmed that each C-propeptide chain has five domains; from N- to C-terminal these are the telopeptide attachment sequence, the interchain disulfide bonding sequence, G1, Lf-Lr (Lfr), and G2. Assuming that the prior literature on the chain synthesis and folding is correct, then folding and interchain association begin in the C-propeptide and proceed in the N-terminal direction. The presence of the most C-terminal sequence of C-pro α2(I), designated here as the Lr sequence, is crucial for the chain selection and insertion of the α2(I) chain into the heterotrimer (*4*, *13*, *22*). As the nascent chains are elongated and passed through the ER membrane into the ER cisternal space, the very long α chains do not associate with each other but do have many interactions with the ER-resident posttranslational chain modifying enzymes and chaperone proteins such as HSP47 (*46*, *47*). During this period each chain segment begins to search for its most stable conformation and assume local folded conformations. Since the Gly-Xaa-Yaa domains, C-telopeptides, N-attachment, and interchain interaction sequences all seem to favor extended chain conformations prior to trimerization, we propose that the G1, Lfr, and G2 domains are the first to assume defined folds. In the homotrimer the α1 Lfr-G2 domains do not strongly interact with each other, suggesting that trimerization begins in the G1 domains (Figures 2 and 3). This is consistent with the Doyle and Smith observation that removal of the α1 Lr sequences does not inhibit trimerization (*22*). On the other hand, the folding of G2α2 which turns the C-terminal α2 Lr back to meet G1α2 so that it can form antiparallel *β*-segments with α2 Lf and allow the intrachain 84−245 disulfide bond formation connecting the C-terminal structure to G1α2 is crucial, as is the formation of the G2α2 153−198 disulfide bond. As shown in Figures 2 and 3, the G2α2-Lfr junction strongly interacts with the similar region on one α1 chain, the one which will become the $\alpha1_2$ chain in the final structure. Thus we suggest that this initial dimer formation is critical in setting the ultimate $\alpha1_1$-α2-$\alpha1_2$ helix chain stagger within the heterotrimer. The $\alpha1_1$ chain then associates with the dimer at the G1 or Lfr domain near G2, and the trimer is formed. Alignment and disulfide linking then proceed through the trimer in the interchain disulfide linkage segment.

Although the arguments of McAlinden et al. (*26*) about the trimerization zone being at the most N-terminal part of the C-propeptide are compelling in terms of the coiled-coil argument and agreement for type II and type III homotrimers, this does not seem to be possible for the type I heterotrimer since that region of the α2 chain has proline residues that would prevent α-helical coiled-coil formation. McAlinden et al. (*26*) recognized the presence of the prolines in the N-terminal region of the C-pro α2 chain and suggested that this was the basis for the difficulty of forming pro α2(I) homotrimers. However, our modeling data for the α1(I) homotrimer show no evidence of an α-helical-based coiled-coil structure preferentially forming in type I collagen. Figure 12 shows the telopeptide and putative trimerization zone C-propeptide sequences for the α1 and α2 chains. There are clearly two distinct Pro-rich sequence regions where there are difficulties in driving the three-chain assembled molecule through into an in register triple helix, one in the α1 C-telopeptides and the other in the α2 C-pro peptide. As illustrated in Figure 6A, and at higher resolution in Figure 6B,C, in the homotrimer, the three α1 chains of the C-telopeptide have their poorest folding potential in the C-telo Pro-rich sequence. The openness of that region, however, is important in the intermolecular assembly of the processed collagen into fibrils when the C-propeptide is cleaved by the C-proteinase, as suggested in the study of the free telopeptide structure (*30*). In the most stable conformation of the heterotrimer, the sequence at the junction between C-pro and C-telo is for a tight $\alpha1_1$-$\alpha1_2$ alignment

with the α2 Pro-rich sequence looped out, while within the C-telo assembled heterotrimer, the α2-α1_2 stagger provides a good dimer alignment while the α1_1 Pro-rich sequence is looped out. If, as suggested by the greater intrinsic stability of the helical domain of the [α1]$_3$ homotrimer (*44*) that structure is energetically favored, then the kinetics of folding through the C-telo-C-propeptide junction, based on the presence of the α2 chain, may be the basis for the preferential heterotrimer formation. Simply studying the trimerization and folding in the C-propeptide alone is not sufficient to explain the problem of folding the entire collagen molecule; one must consider the intact C-telopeptide. The model proposed here is compatible with the electron microscopy of segment-long-spacing crystallites of type I procollagen, explicitly showing the attachment of the folded propeptide with a short linear extension from the C-terminal end of the triple helix (*48*), yielding an overall propeptide length of 150−170 Å.

The α2 chain and the α1_1-α2-α1_2 chain stagger within the triple-helix segment are clearly key elements in collagen assembly, and the differences between the α2 chain and α1 chain sequences within the C-propeptide and C-telopeptide domains appear to determine the order of chain assembly. These considerations all suggest that it is the selective second-order dimerization reaction between a G2α1_2 and G2α2 domain that initiates assembly while trimerization at the G1 domain registers the three chains. The junctions between the Lfr and G2 domains are probably an equally important part of the initial chain association interactions, but that remains to be explored in greater detail. From these models, one can now examine the effects of various C-propeptide mutations more fully.

ACKNOWLEDGMENT

We thank Olaseni Solomon Oduwaye for technical assistance in preparing the maltose binding protein constructs and for help with the dynamic light scattering.

REFERENCES

1. Doege, K. J., and Fessler, J. H. (1986) Folding of carboxyl domain and assembly of procollagen I, *J. Biol. Chem. 261*, 8924−8934.
2. Veis, A., Leibovich, S. J., Evans, J., and Kirk, T. Z. (1985) Supramolecular assemblies of mRNA direct the coordinated synthesis of type I procollagen chains, *Proc. Natl. Acad. Sci. U.S.A. 82*, 3693−3697.
3. Veis, A., and Kirk, T. Z. (1989) The coordinate synthesis and cotranslational assembly of type I procollagen, *J. Biol. Chem. 264*, 3884−3889.
4. Hu, G., Tylzanowski, P., Inoue, H., and Veis, A. (1995) Relationships between translation of pro α1(I) and pro α2(I) mRNAs during synthesis of the type I procollagen heterotrimer, *J. Cell. Biochem. 59*, 214−234.
5. Gura, T., Hu, G., and Veis, A. (1996) Posttranscriptional aspects of the biosynthesis of type I collagen pro-alpha chains. The effects of posttranslational modifications on synthesis pauses during the elongation of the pro α1(I) chain, *J. Cell. Biochem. 61*, 194−215.
6. Beck, K., Boswell, B. A., Ridgway, C. C., and Bachinger, H. P. (1996) Triple helix formation of procollagen type I can occur at the rough endoplasmic membrane, *J. Biol. Chem. 271*, 21566−21573.
7. Bateman, J. F., Lamande, S. R., Dahl, H. H., Chan, D., Mascara, T., and Cole, W. G. (1989) A frameshift mutation in results in a truncated non-functional carboxyl-terminal pro alpha 1(I) propeptide of type I collagen in osteogenesis imperfecta, *J. Biol. Chem. 264*, 10960−10964.
8. Chessler, S. D., Wallis, G. A., and Byers, P. H. (1993) Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of type I collagen result in defective chain association and produce lethal osteogenesis imperfecta, *J. Biol. Chem. 268*, 18218−18215.
9. Chessler, S. D., and Byers, P. H. (1993) BiP binds type I procollagen alpha chains with mutations in the carboxyl-terminal propeptide synthesized by cells from patients with osteogenesis imperfecta, *J. Biol. Chem. 268*, 18226−18233.
10. Cole, W. G., Chow, C. W., Bateman, J. F., and Sillence, D. O. (1996) The phenotypic features of osteogenesis imperfecta resulting from a mutation of the carboxyl-terminal pro alpha 1(I) propeptide that impairs the assembly of type I procollagen and the extracellular matrix, *J. Med. Genet. 33*, 965−967.
11. Pace, J. M., Kuslich, C. D., Willing, M. C., and Byers, P. H. (2001) Disruption of one intra-chain disulphide bond in the carboxyl-terminal propeptide of the proalpha 1(I) chain of type I procollagen permits slow assembly and secretion of overmodified, but stable procollagen trimers and results in mild osteogenesis imperfecta, *J. Med. Genet. 38*, 443−449.
12. Kuznetsova, N. V., McBride, D. J., Jr., and Leikin, S. (2003) Changes in thermal stability and microunfolding pattern of collagen helix resulting from the loss of alpha2(I) chain in osteogenesis imperfecta murine, *J. Mol. Biol. 33*, 191−200.
13. Alvares, K., Siddiqui, F., Malone, J. P., and Veis, A. (1999) Assembly of the type I procollagen molecule: Selectivity of the interactions between the pro-α1(I) and pro-α2(I)-carboxyl propeptides, *Biochemistry 38*, 5401−5411.
14. Koivu, J. (1987) Disulfide bonding as a determinant of the molecular composition of types I, II, and III procollagen, *FEBS Lett. 217*, 216−220.
15. Chuang, C. C., Chen, C. Y., Yang, J. M., Lyu, P. C., and Hwang, J. K. (2003) Relationship between protein structures and disulfide-bonding patterns. *Proteins: Struct., Funct., Bioinf. 53*, 1−5.
16. Sundaramoorthy, M., Meiyappan, M., Todd, P., and Hudson, B. G. (2002) Crystal structure of NC1 domains. Structuralbasis for type iv collagen assembly in basement membranes, *J. Biol. Chem. 277*, 31142−31153.
17. Hohenester, E., Sasaki, T., Olsen, B. R., and Timpl, R. (1998) Crystal structure of the angiogenesis inhibitor endostatin at 1.5 Å resolution, *EMBO J. 17*, 1656−1664.
18. McLaughlin, S. H., and Bulleid, N. J. (1998) Molecular recognition in procollagen chain assembly, *Matrix Biol. 16*, 369−377.
19. Bulleid, N. J., Dalley, J. A., and Lees, J. F. (1997) The C-propeptide domain of procollagen can be replaced with a transmembrane domain without affecting trimer formation or triple helix folding, *EMBO J. 16*, 6694−6701.
20. Myllyharyu, J., Lamberg, A., Notbohm, H., Fietzek, P. P., Pihlajaniemi, T., and Kivirikko, K. I. (1997) Expression of wild-type and modified pro α chains of human type I procollagen in insect cells leads to the formation of stable [α1(I)]$_2$α2(I) collagen heterotrimers and [α1(I)]$_3$ homotrimers but not [α2(I)]$_3$ homotrimers, *J. Biol. Chem. 272*, 21824−21830.
21. Lim, A. L., Doyle, S. A., Balian, G., and Smith, B. D. (1998) Role of the pro-alpha2(I) COOH-terminal region in assembly of type I collagen: truncation of the last 10 amino acid residues of pro-alpha2(I) chain prevents assembly of type I collagen heterotrimer, *J. Cell. Biochem. 71*, 216−232.
22. Doyle, S. A., and Smith, B. D. (1998) Role of the pro-alpha2(I) COOH-terminal region in assembly of type I collagen: disruption of two intramolecular disulfide bonds in pro-alpha2(I) blocks assembly of type I collagen, *J. Cell. Biochem. 71*, 233−242.
23. Hohenester, E., and Engel, J. (2002) Domain structure and organisation in extracellular matrix proteins, *Matrix Biol. 21*, 115−128.
24. Hulmes, D. J. (2002) Building collagen molecules, fibrils, and suprafibrillar structures, *J. Struct. Biol. 137*, 2−10.
25. Engel, J., and Kammerer, R. A. (2000) What are oligomerization domains good for? *Matrix Biol. 19*, 283−288.
26. McAlinden, A., Smith, T. A., Sandell, L. J., Ficheux, D., Parry, D. A., and Hulmes, D. J. (2003) Alpha-helical coiled-coil oligomerization domains are almost ubiquitous in the collagen superfamily, *J. Biol. Chem. 278*, 42200−42207.
27. Veis, A., Alvares, K., and Malone, J. P. (1999) Type I procollagen heterotrimer assembly is linked to subtle differences in the structures of the pro-α1(I) and pro-α2(I)-carboxyl propeptides, *Proc. Indian Acad. Sci. (Chem. Sci.) 111*, 115−120.
28. Bernocco, S., Finet, S., Ebel, C., Eichenberger, D., Mazzorana, M., Farjanel, J., and Hulmes, D. J. (2001) Biophysical characterization of the C-propeptide trimer from human procollagen III reveals a tri-lobed structure, *J. Biol. Chem. 276*, 48930−48936.

29. Malone, J. P., George, A., and Veis, A. (2004) Type I collagen N-telopeptides adopt an ordered structure when docked to their helix receptor during fibrillogenesis, *Proteins: Struct., Funct., Bioinf. 54*, 206−215.
30. Malone, J. P., and Veis, A. (2004) Heterotrimeric type I C-telopeptide conformation as docked to its helix receptor, *Biochemistry 43*, 15358−15366.
31. Dauber-Osguthorpe, P., Maunder, C. M., and Osguthorpe, D. J. (1996) Molecular dynamics: deciphering the data, *J. Comput.-Aided Mol. Des. 10*, 177−185.
32. 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.
33. Takagi, H., Morinaga, Y., Tsuchiya, M., Ikemura, G., and Inouye, M. (1988) *Bio/Technology 6*, 948−950.
34. Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari A. (2002) Crystal structure of the collagen triple helix model $[(Pro\text{-}Pro\text{-}Gly)_{(10)}]_3$, *Protein Sci. 11*, 262−270.
35. Koivu, J. (1987) Identification of disulfide bonds in carboxy-terminal propeptides of human type I procollagen, *FEBS Lett. 212*, 229−232.
36. Koivu, J., and Myllyla, R. (1987) Interchain disulfide bond formation in types I and II procollagen. Evidence for a protein disulfide isomerase catalyzing bond formation, *J. Biol. Chem. 262*, 6159−6164.
37. Pesciotta, D. M., Dickson, L. A., Showalter, A. M., Eikenberry, E. F., de Crombrugghe, B., Fietzek, P. P., and Olsen, B. R. (1981) Primary structure of the carbohydrate-containing regions of the carboxyl propeptides of type I procollagen, *FEBS Lett. 125*, 170−174.
38. Marti, T., Kosen, P. A., Honory, K., Franzblau, C., Schmid, K., Van Halbeek, H., Gerwig, G. J., and Vliegenthart, J. F. (1984) The structure of the carbohydrate units of the carboxyl-terminal peptide of procollagen as elucidated by 500 MHz ^{1}H NMR spectroscopy, *Biochim. Biophys. Acta 799*, 305−312.
39. Lamande, S. R., and Bateman, J. F. (1995) The type I pro alpha 1(I) COOH-terminal propeptide N-linked oligosaccharide. Functional analysis by site-directed mutagenesis, *J. Biol. Chem. 270*, 17858−17865.
40. Olsen, B. R., Cuzman, N. A., Engel, J., Condit, C., and Aase, S. (1977) Purification and characterization of a peptide from the carboxyl-terminal region of chick tendon procollagen type I, *Biochemistry 16*, 3030−3036.
41. Shilton, B. H., Shuman, H. A., and Mowbray, S. L. (1996) Crystal structures and solution conformations of a dominant negative mutant of *Escherichia coli* maltose-binding protein, *J. Mol. Biol. 264*, 364−376.
42. Spurlino, J. C., Lu, G. Y., and Quiocho, F. A. (1991) The 2.3 Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis, *J. Biol. Chem. 266*, 5202−5219.
43. Ramirez, F. (1989) Organization and evolution of the fibrillar collagen genes, in *Collagen* (Olsen, B. R., and Nimni, M. E., Eds.) Vol. 4, pp 21−30, CRC Press, Boca Raton, FL.
44. Leikina, E., Mertts, M. V., Kuznetsova, N., and Leikin, S. (2002) Type I collagen is thermally unstable at body temperature, *Proc. Natl. Acad. Sci. U.S.A. 99*, 1314−1318.
45. Persikov, A. V., and Brodsky, B. (2002) Unstable molecules form stable tissues, *Proc. Natl. Acad. Sci. U.S.A. 99*, 1101−1103.
46. Sauk, J. J., Smith, T., Norris, K., and Ferreira, L. (1994) Hsp47 and the translation-translocation machinery cooperate in the production of alpha 1(I) chains of type I procollagen, *J. Biol. Chem. 269*, 3941−3946.
47. Ferreira, L. R., Norris, K., Smith, T., Hebert, C., and Sauk, J. J. (1994) Association of Hsp47, Grp78, and Grp94 with procollagen supports the successive or coupled action of molecular chaperones, *J. Cell. Biochem. 56*, 518−526.
48. Bruns, R. R., Hulmes, D. J. S., Therrien, S. F., and Gross, J. (1979) Procollagen segment-long-spacing crystallites: Their role in collagen fibrillogenesis, *Proc. Natl. Acad. Sci. U.S.A. 76*, 313−317.

BI0508338