Heterotrimeric Type I Collagen C-Telopeptide Conformation As Docked to Its Helix Receptor†

James P. Malone and Arthur Veis*

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

Received August 6, 2004; Revised Manuscript Received September 24, 2004

ABSTRACT: The amino (N-telo) and carboxyl (C-telo) telopeptides of type I collagen play crucial roles, *in vivo* and *in vitro*, in the assembly of collagen fibrils, regulating the axial alignment of the molecules within a fibril, azimuthal orientations of neighboring molecules, and cross-link formation. High-resolution structures of the telopeptides are not available from X-ray diffraction studies, but computational methods permitted prediction of the N-telo structure within a fibril. Here, using a suite of molecular modeling software, the more complex heterotrimeric C-telo of human type I collagen has been built from the correct sequences and energy minimized and the energy minimum confirmed by molecular dynamics. The receptor triple helix was modeled on the basis of the Protein Data Bank coordinates of a collagen-like sequence. Docking of the heterotrimeric C-telopeptide to its receptor showed that hydrophobic interactions involving the short α2 C-telopeptide are crucial determinants of its azimuthal orientation within the docked structure. A docked C-telo can interact with only one neighboring helix. The two α1(I) C-telo chains in the α1(A)−α2−α1(B) chain stagger do not have identical docked conformations, and one of the α1(I) C-telo chains appears to be favored for formation of a cross-link between its K^{16C} and a helix K87. Prior studies showed that a docked N-telopeptide can interact with two adjacent collagen monomers, forming a tightly packed region. A recent X-ray analysis showed the N- and C-telo regions pack differently, with the C-telo region being less densely packed than N-telo regions. This difference between N- and C-telopeptide docked structures demonstrates how unique and specific packing can occur in the fibril at each boundary of the type I collagen gap region.

Shortly after a newly synthesized type I procollagen molecule is secreted from the cell into the extracellular matrix, the propeptides are excised to produce the fibril-forming collagen molecule (COL1) with its amino (N) and carboxyl (C) telopeptides exposed. These short telopeptide domains participate in the interactions that bring the molecules into ordered fibrils with the appropriate quarter-stagger registration in the Hodge−Petruska model, and ultimately also participate directly in intermolecular cross-link formation. We have recently modeled the conformation, packing, and cross-linked structures of the N-telopeptides (*1*) of the COL1 heterotrimer. That study brought out three very important points. First, when the individual α1 or α2 N-telopeptides were considered, their chain structures were essentially random, as Jones and Miller (*2*) had suggested. Second, in sharp contrast, when the telopeptides were docked to their helix domain receptors on an adjacent molecule, very ordered and specific conformations were created. Third, the receptor region triple-helix conformation was also modulated in a reciprocal fashion by the docking interaction. As had been pointed out long ago (*3*, *4*) in the first detailed studies of this problem, the triple-helix sequences in the cross-linking regions, relatively poor in Gly-Pro-Pro and Gly-Pro-Hyp triplets, were likely to be less conformationally stable than other parts of the COL1 triple helix in the collagen monomer. The study by Malone *et al.* (*1*) showed indeed that the docking of the N-telopeptide to its triple-helical receptor domain induced particularly stable but modified helical conformations of the triple helix when associated with the telopeptides, varying with the nature of the cross-links that are formed.

The C-telopeptide helix docking problem is formally similar to that of N-telopeptide docking, but there are subtle differences in both telopeptide and helix domain sequences that frustrated the early attempts to model the C-telopeptide structure and interactions (*5*−*9*). These small differences have very great consequences in that the cross-links involved with the C-telopeptide and its helix receptor are different from those of the N-telopeptide (*10*), and the supramolecular packing in the C-telo domain is different from that in the N-telo domain (*11*).

In this work, the C-telopeptide conformation as docked to its helix receptor domain has been modeled using an approach similar to that used for the N-telopeptide. That is, the individual human α1(I) and α2(I) C-telopeptide sequences were modeled separately using energy minimization computations. The individual chains were then joined to the most C-terminal triple-helical segment, using the real human COL1 sequences to register the chains in the α1−α2−α1 chain stagger order found to be the best fit for N-telopeptide

† This work has been supported by National Institute of Arthritis and Musculoskeletal and Skin Diseases Grant AR-013921 (A.V.).

* To whom correspondence should be addressed: Department of Cell and Molecular Biology, Feinberg School of Medicine, Northwestern University, 303 E. Chicago Ave., Chicago, IL 60611. Phone: (312) 503-1355. Fax: (312) 503-2544. E-mail: aveis@northwestern.edu.

10.1021/bi048304b CCC: $27.50
Published on Web 11/18/2004

modeling (*1*). The in-register three-chain telopeptide unit was then energy-minimized. Next, the triple-helix receptor sequence surrounding the cross-linking receptor lysine, α1-K87, was modeled, using the correct human sequences for all three triple-helical chains. The three-chain C-telo structure was then docked in the correct registration and orientation to the helix domain, and the entire docked conformation was determined. A distinct structure, different from the N-telopeptide conformation but compatible with the X-ray data-based predictions of Orgel *et al.* (*12*) for the C-telopeptide domain, has been determined.

HUMAN COL1 C-TELOPEPTIDES

The sequences of the human COL1 C-telopeptides, according to the latest SWISS-PROT data bank, are as follows:

Accession #P02452 α1 (I) **SAGFDFSFLPQPPQE K^{16} AHDGGRYYRA^{26}**

Accession #P08123 α2 (I) GGGYDGYDGD----------FYRA^{14}

The only homology between the α1 and α2 C-telopeptides is in the final four residues that comprise part of the recognition site for the action of the C-proteinase that cleaves the C-propeptide from the C-telopeptide. NMR studies of the α2(I) C-telopeptide in solution showed it to have a random extended structure throughout (*13*). Similar NMR studies had shown that the isolated α1(I) C-telopeptide was also essentially free of any defined structure (*14*). Notably, although the amino-terminal sequence of the α1(I) C-telopeptide has a concentration of bulky hydrophobic phenylalanine residues, these are followed by a proline-rich sequence that kinks the extended chain. Only the α1 C-telopeptides have a Lys residue, K^{16C}, which can be used in cross-linking. This is distinctly different from the case for the N-telopeptides, in which all three chains can participate in cross-link formation.

MODELING COMPUTATIONS AND BASIC ASSUMPTIONS

General Predictions. The structure of the free α1(I) C-telopeptide was examined first by general predictive algorithms using the PROTEAN module of Lasergene/DNAstar. The α-helix, *β*-structure, and turn content was estimated using the Garnier−Robson and Chou−Fasman algorithms. In agreement with the NMR data cited above, these generalized predictions indicated that the free α1(I) C-telopeptide sequence was unlikely to adopt any extended ordered α-helical conformations or *β*-structures (Figure 1).

Similar results, with some preference for *β*-structures, were obtained for the free α2(I) C-telopeptide. Each of the algorithms that is shown confirms the general feature that there are four distinct domains of different character in the α1 C-telopeptide sequence. In Figure 1, the linear sequence schematic along the bottom of the α1(I) C-telo plot, with phenylalanine-rich, proline-rich, hydrophilic lysine-containing, and tyrosine-rich sequence segments, emphasizes the nature of the domains in the α1(I) telopeptide. The sequence of the α2(I) C-telo, also displayed, is distinctly different.

Although we can examine the properties of the individual chains as described above, it is important to emphasize that the C-telopeptide chains never exist in the separate, free state in the collagen I molecule. They are properly aligned and

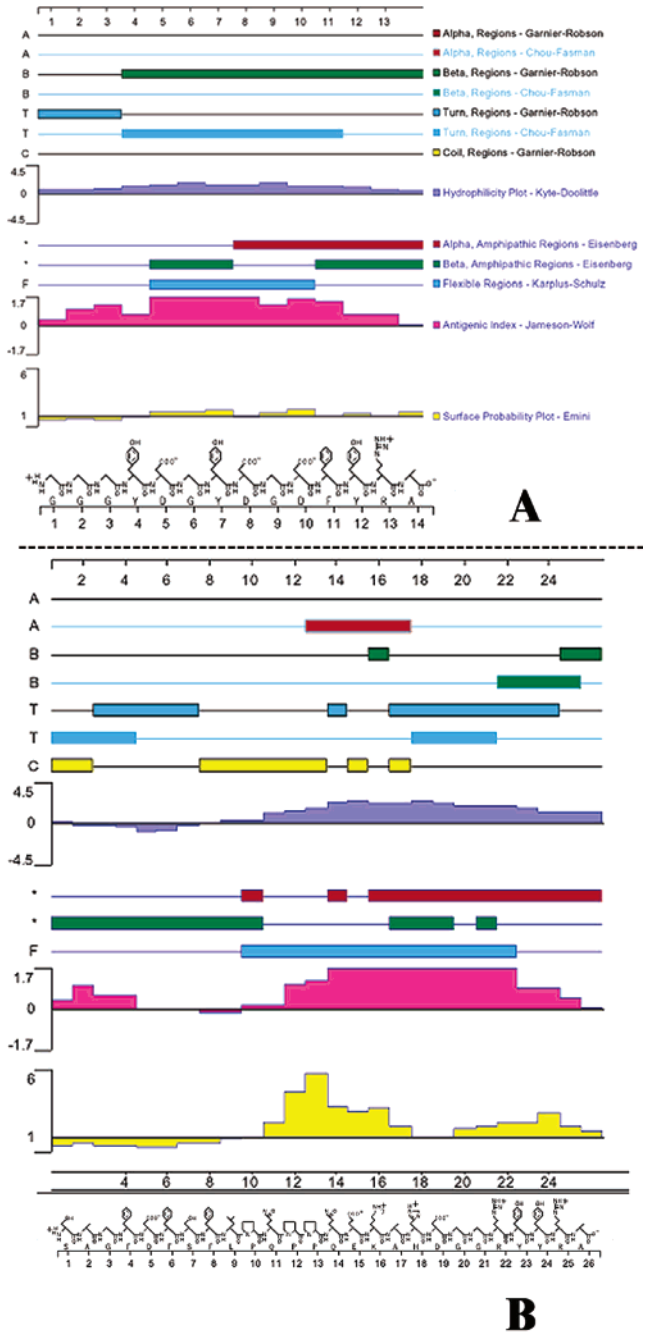

FIGURE 1: Conformation and properties of the COL1 C-telopeptides as predicted using the PROTEAN module of Lasergene/DNAstar: (A) α2(I) and (B) α1(I). The color coding is the same in panels A and B.

assembled in some structure as the connecting link between the C-propeptide and the triple helix. Given that the formation and folding of the complete procollagen molecule proceed from the C-terminus to the N-terminus, the C-telopeptide has two distinct boundary conditions. There is the chain alignment determined by the trimerization zone within the C-propeptide (*15*) and the well-defined, very stable triple-helical C-terminal segment of Gly-Pro-Hyp repeats. The connecting region between the C-telopeptide and C-propeptide has been examined in connection with our detailed study of the C-propeptide structure. That work will be reported later. From that work, it is evident that when the C-propeptidase cleaves off the C-propeptide, the structural constraint at the new telopeptide C-terminus is released, and

this is especially important since, as shown above, the three telopeptide chains are not equally long. There is thus a rearrangement in telopeptide conformation, but it is anchored and registered at the N-terminal, triple-helical end. As described below, we have built the C-telopeptide model in the N-terminal to C-terminal direction in steps leading to the heterotrimeric structure.

A second important concept for the modeling is that the docked structure must be concordant with the formation of the well-documented cross-links involving K^{16C} of the C-telopeptide and its triple-helix receptor region cross-link partner at K87 on an adjacent molecule. Finally, we recognized that the helix receptor region has a low Pro and Hyp content compared to the "ideal" triple-helix structure. Thus, it too had to be energy-minimized and its domain structure determined, and then reminimized during the docking process.

Energy Minimization and Structure Prediction. Using a molecular modeling suite of programs from ACCELRYS/MSI, including BIOPOLYMER, DISCOVER, and INSIGHT II, an SGI Octane R12000 SE computer was used for all computations. Several distinct stages were used for the conformational analyses. First, the human C-telopeptide sequences provided above were separately loaded into the minimization program in an extended chain conformation, and a probable structure was determined for each. Next, a short triple-helical sequence was built corresponding to the correct C-terminal sequences of the human heterotrimeric type I collagen triple helix in the α1−α2−α1 chain stagger. 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 [entry 1k6f (*16*)]. The correct amino acids were substituted, and the modified conformation was determined. The minimized individual telopeptide chains were then grafted to the C-terminus of the computed triple helix in the proper stagger to register the chains. The total three-chain structure was then energy-minimized to reflect the interactions between the telopeptide sequences in the intact three-chain molecule. The system was solvated with a 3 Å shell of water (data not shown). The resulting conformation remained essentially unchanged. Similarly, the putative triple-helical region containing the binding site for the telopeptide interactions, helix residues 72−102 in each chain, was modeled, beginning with the coordinates for the pure "$(P\text{-}P\text{-}G)_n$" triple helix. Then the real sequences for each chain of the human type I collagen were inserted by replacement. The triple-helical segment was then energy-minimized.

The two three-chain molecules, the C-terminal helix−telopeptide extensions, and the triple-helical receptor molecule were brought together so that the interaction of at least one Lys-K87 of the helical segment with one of the α1 C-telo K^{16C} residues was possible. The effect of water was not considered at this point because the data set became too large and computationally expensive. However, results of solvation of the individual parts suggested that solvation of the complete docked structure would not change in a major way. The structure of the composite docked structure was then determined by an extended minimization procedure. Finally, 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. In all the computations mentioned above, the consistent valence force field (CVFF) algorithms (*17*) were used to minimize the total energy. Reiterations of the derived structure were continued until the root-mean-square deviation (rmsd) was less than 0.0001. All of the parameters of the energy-minimized structures are made available as PDB files.

RESULTS

Isolated α1(I) C-Telopeptide Sequences. Although only a short sequence, the α1(I) C-telopeptide can be thought to have four distinct segments as emphasized in Figure 1. In the first segment at the amino terminus, the Phe residues at positions 4, 6, and 8 in a trans-extended chain conformation all fall on the same side of the chain, while the intervening polar side chains of Asp and Ser are on the opposite side. The second segment, with Pro residues at positions 10, 12, and 13, changes the direction of the chain from trans-extended and forms a loop. The Lys K^{16C} at the C-end of the loop is thus directed out from the strict N → C direction, as depicted in Figure 2.

Assembled Heterotrimeric C-Telopeptide. The kinked structure of the individual α1(I) C-telopeptide chain is retained in the heterotrimeric three-chain structure, including the α2(I) C-telopeptide chain as shown in the energy-minimized form in Figure 3A.

Within the shorter α2(I) chain domain, there are hydrophobic interactions between the three chains in the segment 1 region circled at the left (Figure 3A). The circled region at the right is a second region of hydrophobic interaction between the two α1(I) C-telopeptide segment 4 domains. The charged residues in segment 3, including α1(I) $Lys16^C$, have much more conformational freedom. Note, in Figure 3B, that the probable *β*-sheet structure (yellow arrow along the two telopeptide chains) that forms between the α2(I) heterotrimer telopeptide chain and one α1 chain (α1B). In the absence of the α2 chain segment, the assembled C-telopeptide (Figure 3C) has a shortened and more globular arrangement and the three Lys residues are in the highly folded and kinked region. Also note that the proline-rich triple-helix zone of the homotrimer, adjacent to the C-telopeptide, forms a more regular triple-helical structure than the heterotrimer.

K^{16C} is the essential component of segment 3 in the α1(I) C-propeptide, with its potential involvement in chain interaction and cross-linking. The final segment, segment 4 in both the α1 and α2 chains, is the bridge to the C-propeptide domain.

The importance of the short α2(I) C-telopeptide in the compound structure is evident in two ways. First, the amino terminus of the α2(I) C-telopeptide has a Gly-Gly hinge region, followed by a pair of Tyr residues, Y^4 and Y^7, that are positioned so that they can interact hydrophobically with the F^4-F^6-F^8 patches on the α1(I) C-telopeptides, creating an energetically stable three-chain assembly extended in the same direction as the triple-helix axis. Second, the short α2(I) C-telopeptide is devoid of sequences corresponding to segments 2 and 3 of the α1(I) C-telopeptide. However, the α2 C-terminal segment homologous to α1(I) C-telopeptide segment 4 appears to be in close association with α1(I) C-telopeptide segment 2 before cleavage of all three chains

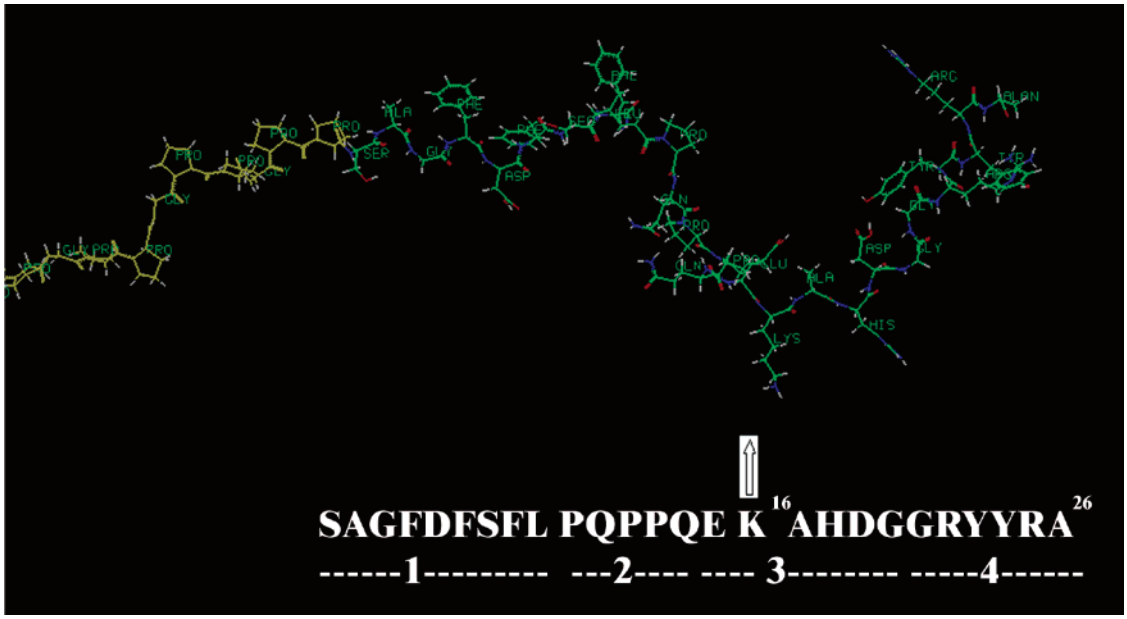

FIGURE 2: Energy-minimized structure of the sequence of the isolated human α1(I) C-telopeptide [SWISS-PROT entry CA11_HUMAN (P02452)] with an rmsd of less than 0.0001. The residues in yellow form the C-terminal portion of the triple-helix region. Numbers 1 (Phe-rich), 2 (Pro-rich), 3 (Lys-containing), and 4 (conserved propeptide-linking) denote the segments of distinct character. Note the location of Lys16 at the apex of the curved structure of segment 3.

of the C-propeptide with the C-proteinase, and α1(I) C-telopeptide segment 3 containing the crucial cross-linking Lys K^{16C} is open to intermolecular interaction.

Assembled α1(I) C-Telopeptide Homotrimer. Once the importance of the α2(I) C-telopeptide became evident, we modeled the same region for the structure of the α1(I) C-telopeptide homotrimer. The structure and net stability of the α1(I) C-telopeptide homotrimer (Figure 3C) is quite distinctly different from that of the heterotrimer depicted in Figure 3A. The computed structure of the C-telopeptide heterotrimer is repeated in Figure 3B, but rotated axially from its representation in Figure 3A so that the strong interaction of the α2 C-telo amino-terminal sequence with only one (not both) of the α1 chains is emphasized. If we consider the helix structure to be of the α1(A)−α2−α1(B) form, then the main helix interaction is with the α1(B) chain. This interaction is not possible in the homotrimer (Figure 3C), in which it is evident that the telopeptide forms a more globular but open domain in which the potential cross-linking Lys residues are located.

Triple-Helix Telopeptide Receptor Docking Region. The receptor region for cross-linking and docking of the C-telopeptide is the sequence surrounding the lysines at residue 87 in each chain (residues 72−102):

α1: Q GAR GLP* GTA GLP* GMK^{87} GHR GFS GLD GAK GDA

α2: G GAR GFP* GTP* GLP* G FK^{87} GIR GHN GLD GLK GQP .

While Pro or Hyp (P*) residues are present (*18*), all are in the third position of each triplet sequence; that is, there are no chain stiffening P-P or P-P* sequences. The energy-minimized structure for the Pro sequences (Figure 4) shows that the helix is less rigid than a $(GPP)_N$ helix, and has a definite curvature and relative looseness of packing near the crucial K87 docking and cross-linking sites. The hydroxylation of these three position sequences, and K87, is not likely to reduce the chain flexibility or affect the docking of the telopeptide. Underhydroxylated or unhydroxylated collagens form fibrils with normal packing, albeit at temperatures lower than those of the fully hydroxylated collagens.

Composite Docked Structures. With the data relating to the structures depicted in Figures 3 and 4 available, the two were oriented in the same N-terminal to C-terminal direction and then brought close together in approximate longitudinal register. The axial alignment and azimuthal orientation were then allowed to develop during an extended docking calculation, which required many thousands of reiterations but finally yielded the structure shown in Figure 5.

It appears that segment 1 of the α1(I) telopeptide, where the α2 chain interaction with the α1 F^4-F^6-F^8 sequence forms a stable structure, binds to the triple-helix region and determines the relative azimuthal orientation of the complex. The helix conformation changes with the reciprocal interaction, slightly unwinding that portion of the docked triple helix. As a result of this α2 chain effect, the orientations of the folded and charged regions 2 and 3 of the α1(A) and α1(B) telopeptides are rotated such that they are not in equivalent juxtaposition with the helix receptor domain. The Lys K^{16C} residues in both α1 telopeptide chains are docked in a registration position potentially allowing them to form cross-links with one of the three helix Lys K87 residues. However, it appears that the α1(B) Lys K^{16C} is more favorably positioned to form a cross-link with either the helix α2 K87 or one of the α1 K87 residues. It is important to recognize that in the docked form, the two α1 C-telopeptide K^{16C} residues are on the surface, readily accessible to the enzymes that are necessary for the oxidative deamination required to initiate the cross-linking reactions.

Another facet of the situation is illustrated in Figure 6, in which the two α1(I) C-telopeptide [α1(A) and α1(B)] docked structures have been separated but kept in their docked chain conformations and in the same orientation as in Figure 5A. Even though both chains have the same sequences, their structures are unique.

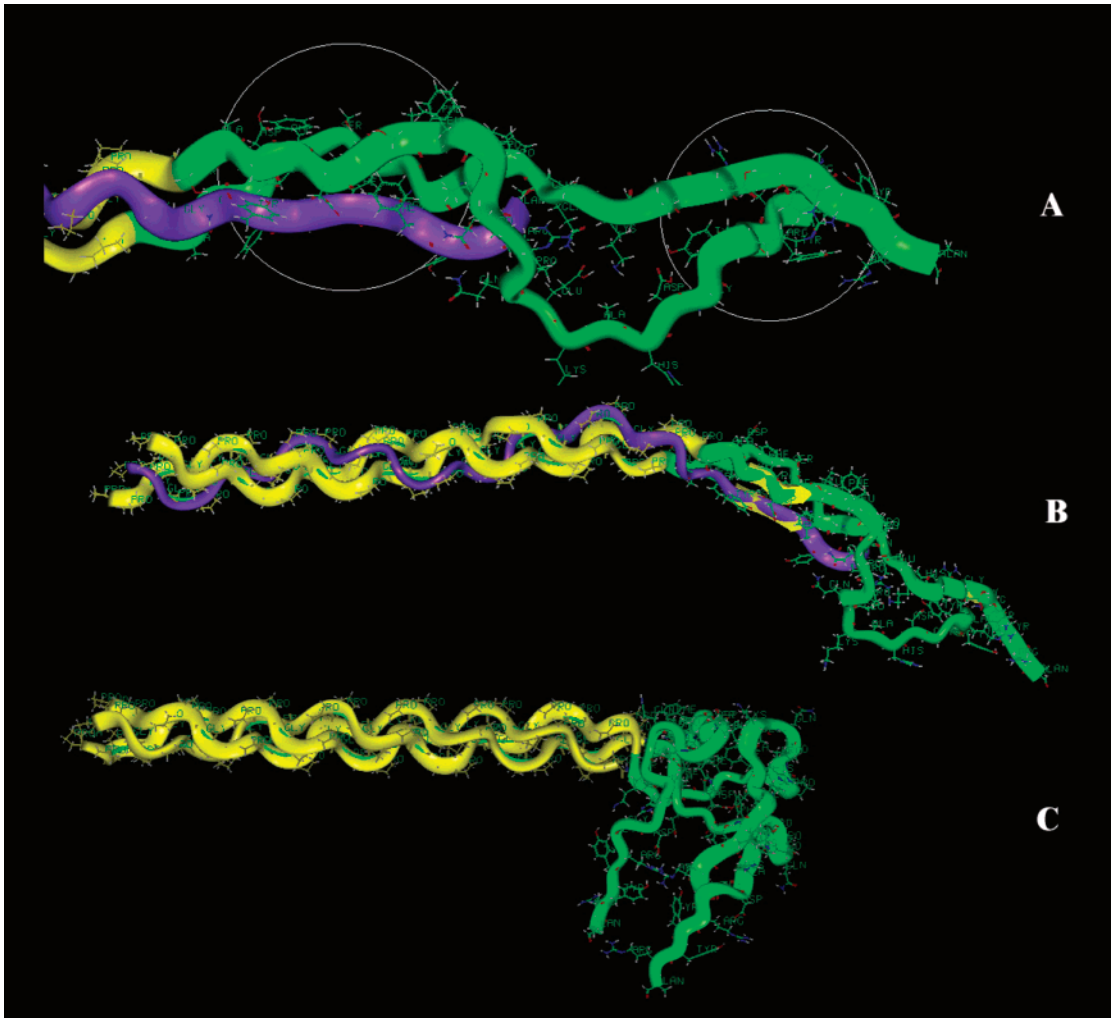

FIGURE 3: Energy-minimized structures of the assembled type I collagen C-telopeptide heterotrimer and the α1 C-telopeptide homotrimer computed using SWISS-PROT entries CA11_HUMAN (P02452) and CA21_HUMAN (P08123) using an α1−α2−α1 stagger. The α1(I) chains of the collagen triple helix are depicted in yellow on the left, changing to green in the telopeptide portion. The α2 chain is in purple in both regions. (A and B) Energy-minimized structure of the heterotrimer in two different orientations. In the orientation shown in panel B, the yellow arrows along the chains within the telopeptide denote a β-sheet conformation involving portions of the α1(B) and α2 chains. (C) Homotrimer energy-minimized conformation.

The top image shows the α1(B) chain, while the bottom image shows the conformation of the α1(A) chain. This depiction emphasizes the fact that each chain has different neighbors and different conformation-modifying environments. The α1(A) chain is more extended and contains the buried K^{16C}, whereas the α1(B) chain has a sharper kink and contains the more exposed K^{16C}. The sharper kink shortens the docked α1(B) chain in the axial direction. It is important to note that both chains are identical and, if treated as separate units, have the same structure as shown in Figure 2. The differences are due to the α1(A)−α2−α1(B) chain stagger and the resultant interchain interactions.

DISCUSSION

This is the first molecular modeling study addressing the type I collagen C-telopeptide conformation using all three chains of the heterotrimer before and after it docks to its receptor domain. The conformation formed at the telopeptides after propeptide excision is determined by the intrinsic properties of each chain sequence and the interactions with the two adjacent, in-register chain sequences. The most carboxyl end of the type I collagen helix adjacent to the C-telopeptide is a region consisting solely of Gly-Pro-Pro-[Hyp] repeats that create a very stable, tight triple-helix collagenous zone with the chains in register. The most amino-terminal of the nonhelical domains of the C-telopeptides are rich in bulky hydrophobic residues [three Phe residues in α1(I) and two Tyr residues in α2(I)] that act to keep the three chains together via hydrophobic interaction, although their conformations are not that of the triple helix. The two α1(I) C-telopeptides extend beyond the α2(I) C-telopeptide and interact with each other as a two-strand structure seeking its lowest-energy resting conformation. This is important as the C-telopeptide interacts with its helix receptor because only the longer α1(I) chains contain the crucial interactive Lys residues. Docking of the C-telopeptide to its receptor domain needs to place these interactive α1(I) telopeptide Lys residues in a position ready for cross-link formation. Although collagen I molecules can form fibrils with the correct stagger without telopeptides present, the telopeptides direct fibril assembly in a much more timely and efficient manner.

Structure predictive algorithms, based strictly on amino acid sequence, demonstrate that there are different domains

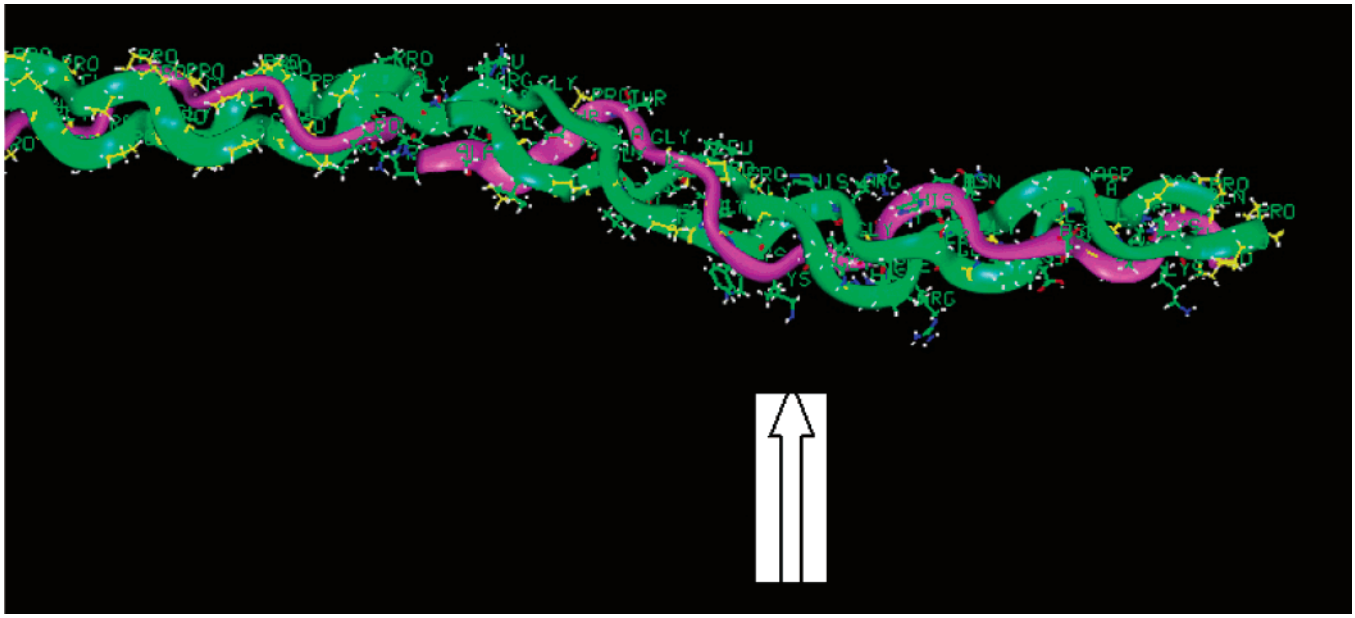

FIGURE 4: Energy-minimized structure of the heterotrimeric type I triple helix computed using the human sequences in the region from residue 72 in each chain through residue 102. The α1−α2−α1 chain registration was used. There is a visible curvature and looseness of packing around the Lys87 receptor region (arrow).

of hydrophobicity, chain flexibility, and polarity present in the α1(I) C-telopeptide (Figure 1). The Pro residues in domain 2 of the α1(I) C-telopeptide bend the chain out of the axial direction (Figure 2), placing the Lys (α1 K^{16C}) that is potentially involved with cross-linking at the apex of the curved structure. When the three chains are assembled together in the α1(A)−α2−α1(B) stagger that provides the best fit at the amino-terminal telopeptide (*1*), the patch of bulky hydrophobic residues contributed by segment 1 in each chain stabilizes an extended structure in the axial direction (Figure 3, top and middle). The α2 chain of the C-telopeptide is much shorter and contributes to the hydrophobic patch but is devoid of the Pro-rich and Lys-containing domains that are looped out in the α1 chain. The two α1(I) chains kink out; however, repulsion between the K^{16C} residues prevents close interaction, and the two chain segments assume different distributions in space with different conformations, placing the K^{16C} residues in different environments. The modeled structure in Figure 3 indicates that E^{15C} of the α1(B) chain may form a salt bridge with α1(A) K^{16C}. The exposed α1(B) K^{16C} most likely prefers to cross-link first to its adjacent monomer receptor lysine, as is seen in the final docked structure views of Figure 5.

When all three chains are α1(I) chains as in the homotrimer illustrated in bottom portion of Figure 3, the three-kinked domain two-Pro-rich segments fold and shorten the whole telopeptide and provide a very different environment for each of the K^{16C} residues in the undocked telopeptide structures. The homotrimer K^{16C} residues are able to form cross-links (*19*), but their orientations may be less directed than in the heterotrimer.

In Figure 4, the helix receptor domain for the C-telopeptide has been modeled and minimized undocked. Cabral *et al*. (*20*) have shown that there is abnormal fibril formation when exon 41, encoding this helix region, is not translated in α1-(I) collagen. The receptor domain is not Pro-rich, and there is neither Pro nor Hyp in the triplet Y position. This allows for considerable relaxation of the triple helix packing which is evident in the computed model that indicates the maximum deviation from the stable triple helix is in the sequence containing K87 in all three chains, the recognition docking site for the C-telopeptide. The flexibility in this site may be needed for docking to occur. As illustrated in Figure 5, the energy-minimized conformation of the docked and assembled telopeptide−receptor complex shows that both partners in the complex have conformations different from that in the free or unassembled state. Both portions change in a reciprocal interaction.

As is obvious in the structure shown in Figure 5, the Pro-rich domain of the telopeptide α1(B) chain is folded over the helix receptor region, aligned with the right-handed superhelix of the receptor domain. This places α1(B) K^{16C} in the proximity of both α2 K87 and only one of the α1 K87s as potential cross-linking partners. When type I collagen is treated with Pronase, the α1 C-telopeptides extending beyond the α2 C-telopeptide are enzymatically removed, but the hydrophobic patch (domain 1) is left intact. If the Pronase-treated collagen is treated with ethylurea, which interferes with hydrophobic interactions, the fibrils dissociate as a result of having no stabilizing C-telopeptide interactions with the receptor collagen monomer. Ninety percent of native collagen (not treated with Pronase) forms normal fibrils after ethylurea treatment, whereas the Pronase- and ethylurea-treated collagen forms tactoid, abnormal structures (*7*). Thus, the interactions of α1 C-telopeptide domains 2−4 are important in establishing both electrostatic and hydrophobic interactions in the telopeptide−receptor interaction.

However, the α2 C-telopeptide portion is also very important in the docking interaction, as suggested in Figure 5. Prockop and Fertala (*21*) demonstrated this experimentally. They studied the thermally induced *in vitro* fibrillogenesis reaction that takes place when pC-collagen is treated with C-proteinase. Synthetic peptides corresponding to α1 and α2 N- and C-telopeptides were added to the fibrillogenesis system. The α2(I) C-telopeptide was found to inhibit fibrillogenesis by 95% when added during the lag phase of

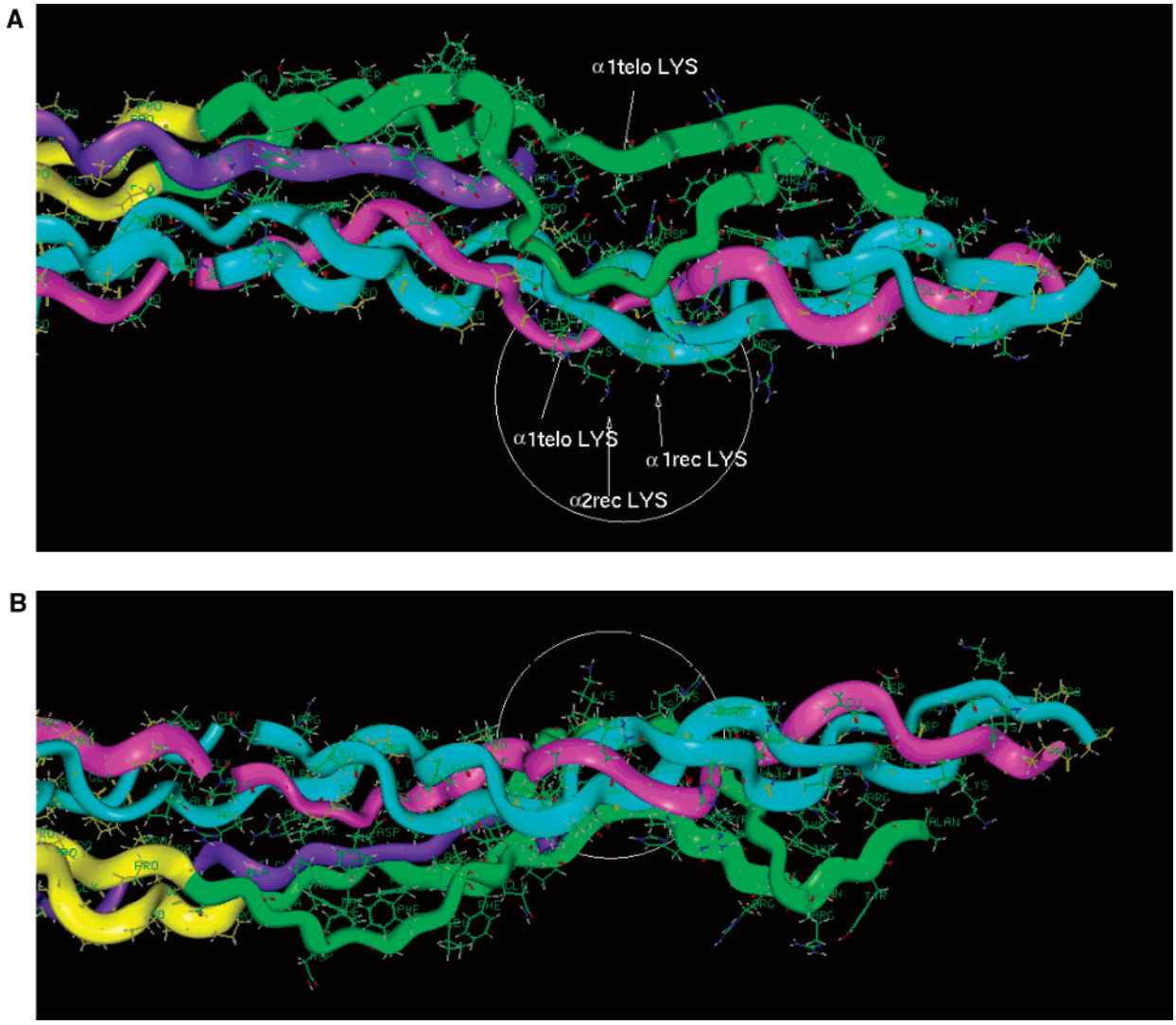

FIGURE 5: Energy-minimized structure of the heterotrimeric type I C-telopeptide docked to its triple-helical receptor domain. Two different views of the docked structure are shown. The helix receptor region α1 chains are colored cyan, and the α2 chain is colored bright purple. The telopeptide α2 chain is colored dark purple, and the α1 chains are colored green. The potential region for interactions for cross-linking is circled. Panels A and B show the same image from opposite sides. The docked α1(I) C-telopeptide chains follow the right-handed superhelix groove of the receptor helix. The α2 chains of both the C-telopeptide and helix receptor interact and set the specificity of axial orientation in the docking assembly.

the reaction, but not later in the assembly. They concluded that "the binding of the α2 C-telopeptide, probably in concert with the α1 C-telopeptide, is critical for early steps in the assembly process such as the formation of a structural nucleus that is essential for further growth of the fibrils." This inhibitory effect of the hydrophobic portion of the free α2(I) C-telopeptide sequence is entirely consistent with the model presented here, if the initial registration of the collagen molecules depends on establishing the quarter-stagger registration of the molecules. The fact that the short α2(I) C-telopeptide peptide sequence also binds to other hydrophobic sequences in the helix region in the Prockop and Fertala (*21*) *in vitro* system experiments does not vitiate the conclusion that the α2(I) C-telopeptide plays a crucial role in the C-telopeptide, helix receptor region registration and assembly.

On the basis of the azimuthal orientation proposed in the model, the two α1(I) C-telopeptide chains only interact directly with the adjacent helix receptor chains after the shorter α2(I) telopeptide ends. The α1 C-telopeptide extensions that fold over the right-handed superhelical groove of the receptor domain are not inhibited in the later fibril assembly processes. The bends in the two α1(I) C-telopeptide chains occur because of two Gly-rich regions that allow flexibility with a central region of three Pro residues producing the bend. The charged residues (EHKD) surrounding the potentially cross-linked K^{16C} allow electrostatic interactions with the receptor domain, followed by a second hinge region of Gly residues. At the end of the α1(I) extensions, multiple tyrosines form a second, small, cohesive hydrophobic zone that keeps the ends of the two chains in contact. The α2(I) chain thus positions the C-telopeptide α1-(B) K^{16C} correctly for cross-link formation with an α1(I) or α2(I) receptor K87. There is only one correct orientation in the C-telopeptide, set by the α2 C-telopeptide chain.

Figure 7 shows the packing density comparing the N-telopeptide with the C-telopeptide derived from recent X-ray diffraction data, confirming the Hulmes−Miller quasi-hexagonal packing in type I collagen (*11*, *22*). In the diagram of the N-telopeptide, segment 1 (the N-telopeptide) of one molecule is able to interact with segment 5 of two adjacent monomers (the N-telopeptide helix receptors surrounding

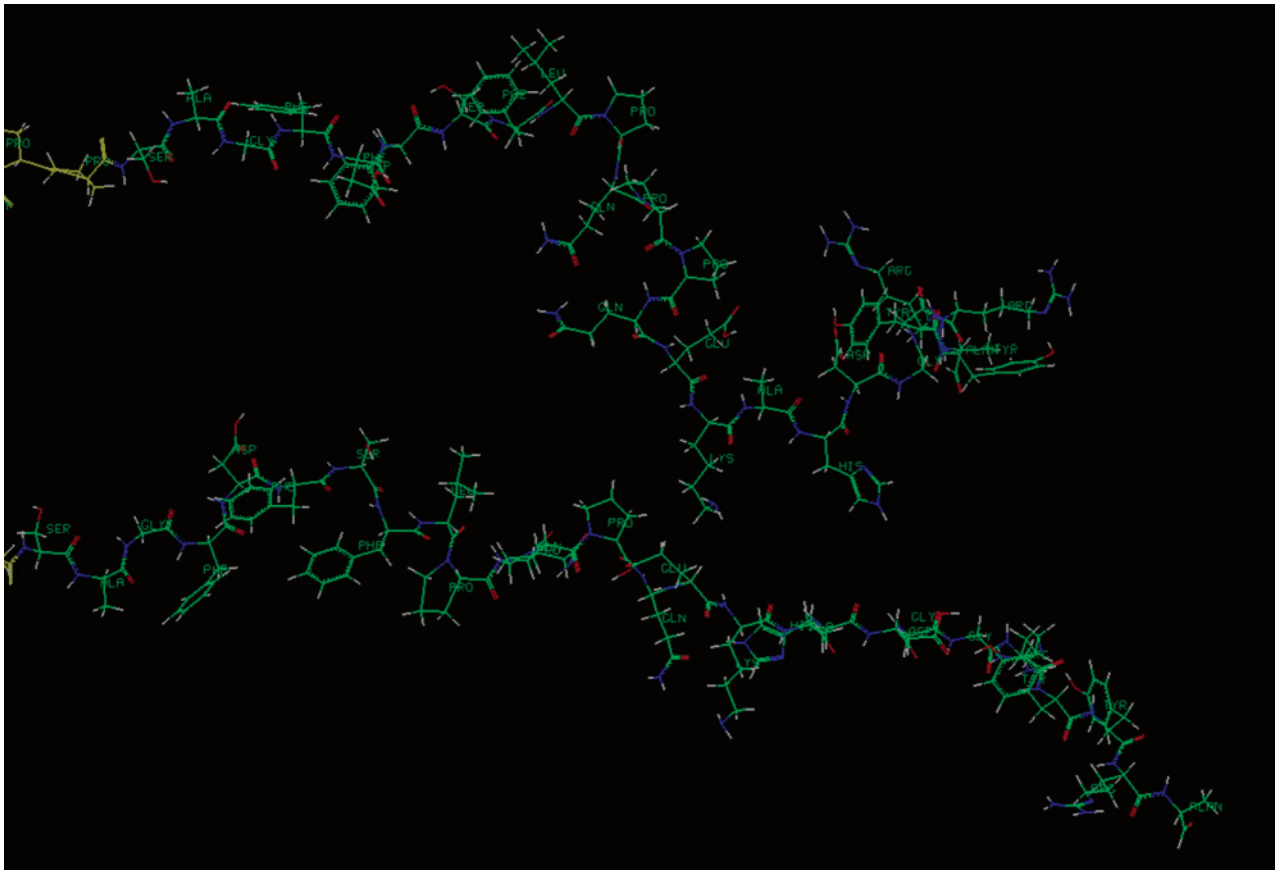

FIGURE 6: Isolated α1(A) and α1(B) telopeptide chains separated but with the conformations computed for the complete telopeptide−helix complex. The top structure is α1(B) and the bottom α1(A). Note how the B chain is kinked and shortened in the helix axial direction as compared to the A chain.

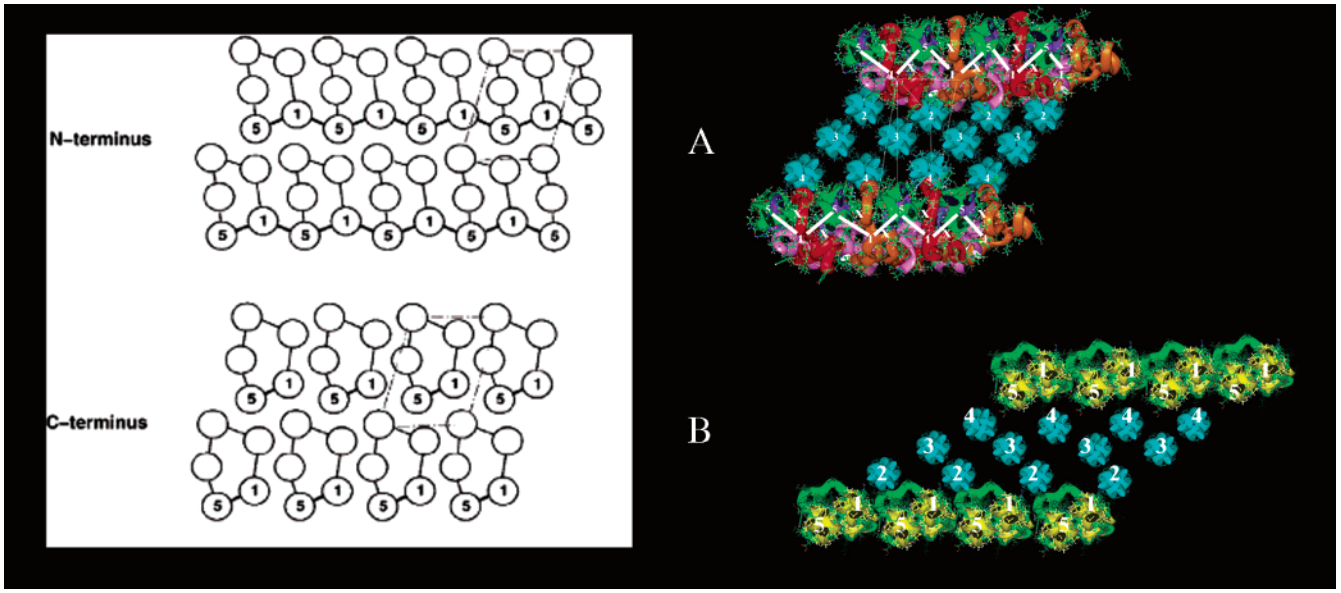

FIGURE 7: Comparison of the fibril cross sections of the packed structures at the amino-terminal and carboxyl-terminal boundaries of the gap region of the collagen fibril. (A) N-Terminal telopeptide and fibril cross structure, as determined by the modeling studies of Malone *et al*. (*1*). Each N-telopeptide can interact with two different collagen monomers, potentially producing transverse cross-linked polymers joining K930 of the α1 CB6 peptide of the helix (*5*) to the K^{7N} of the α1 CB0,1 peptide of the telopeptide domain (*1*) in a (5-1-5-1-5-1-5) array. (B) C-Terminal telopeptide and fibril cross section, as determined in this modeling study. In this case, only one C-telopeptide K^{16C} of α1 CB6 (*5*) is able to interact with the helix K87 (*1*) of only one adjacent collagen monomer in a fibril (5-1). A transverse (5-1-5-1-5-1) cross-linked array is still possible, but the azimuthal orientations and packing density are different. The model agrees well with the differences in packing order seen experimentally in the Hulmes−Miller quasi-hexagonal packing structure between the N-terminus and the C-terminus (*22*). There is tighter packing at the N-terminus and more control of azimuthal orientation by the cross-links that occur.

residue K930), producing a sheet structure and a densely packed zone. This N-telopeptide packing was demonstrated in the model by Malone *et al.* (*1*). The diagram also shows how the C-telopeptide on molecular segment *5* is only able

to interact with one helix receptor molecule at segment 1, producing a zone that is less closely packed than at the N-telopeptide. It appears that there is more control of azimuthal orientation at the N-telopeptide docked region than at the C-telopeptide zone. This becomes important as cross-links form since the types of cross-links are limited by the number of chains that are docked to a monomer. In collagen I, the type of cross-links and maturity level determine tissue destiny (*23*). This is also closely related to Lys hydroxylation in telopeptides and helix receptor domains (*24*). The docking models show clearly that the N- and C-telopeptide regions have different molecular packing and intrafibrillar cross-linking patterns that control the relative azimuthal orientations of molecules in the fibril. The next papers in this series consider the various cross-link structures in detail as well as the C-propeptide structure.

REFERENCES

1. 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.
2. Jones, E. Y., and Miller, A. (1987) Structural models for the N- and C-terminal telopeptide regions of interstitial collagens, *Biopolymers 26*, 463−480.
3. Helseth, D. L., Jr., Lechner, J. H., and Veis, A. (1979) The Role of the Amino-Terminal Extrahelical Region of Type I Collagen in Directing the 4 D Overlap in Fibrillogenesis, *Biopolymers 18*, 3005−3014.
4. Veis, A. (1982) Collagen Fibrillogenesis, *Connect. Tissue Res. 1*, 11−24.
5. Veis, A., and Payne, K. (1988) Collagen Fibrillogenesis, in *Collagen: Chemistry, Biology and Biotechnology* (Nimni, M., Ed.) pp 113−137, CRC Press, Boca Raton, FL.
6. Helseth, D. L., Jr., and Veis, A. (1981) Conformational Studies on the Telopeptides of Collagen. Implications for Intermolecular Interactions, in *Chemistry and Biology of Mineralized Connective Tissues*, pp 85−89, Elsevier North-Holland, New York.
7. Helseth, D. L., Jr., and Veis, A. (1981) Collagen self-assembly in vitro. Differentiating specific telopeptide-dependent interactions using selective enzyme modification and the addition of free amino telopeptide, *J. Biol. Chem. 256*, 7118−7128.
8. Capaldi, M. J., and Chapman, J. A. (1982) The C-terminal extrahelical peptide of type I collagen and its role in fibrillogenesis *in vitro*, *Biopolymers 21*, 2291−2313.
9. Capaldi, M. J., and Chapman, J. A. (1984) The C-terminal extrahelical peptide of type I collagen and its role in fibrillogenesis in vitro: effects of ethylurea, *Biopolymers 23*, 313−323.
10. Otsubo, K., Katz, E. P., Mechanic, G. L., and Yamauchi, M. (1992) Cross-linking connectivity in bone collagen fibrils: the COOH-terminal locus of free aldehyde, *Biochemistry 31*, 396−402.
11. Orgel, J. P., Miller, A., Irving, T. C., Fischetti, R. F., Hammersley, A. P., and Wess, T. J. (2001) The in situ supermolecular structure of type I collagen, *Structure 9*, 1061−1069.
12. Orgel, J. P., and Wess, T. J. (2000) The in situ conformation and axial location of the intermolecular cross-linked non-helical telopeptides of type I collagen, *Structure 8*, 137−142.
13. Liu, X. H., Scott, P. G., Otter, A., and Kotovych, G. (1990) Solution conformation of the type I collagen α-2 chain telopeptides studied by ^{1}H and ^{13}C NMR spectroscopy, *J. Biomol. Struct. Dyn. 8*, 63−80.
14. Otter, A., Scott, P. G., and Kotovych, G. (1988) Type I collagen α-1 chain C-telopeptide: Solution structure determined by 600-MHz proton NMR spectroscopy and implications for its role in collagen fibrillogenesis, *Biochemistry 27*, 3560−3567.
15. McAlinden, A., Smith, T. A., Sandell, L. J., Ficheux, D., Parry, D. A. D., and Hulmes, D. J. S. (2003) α-Helical Coiled-coil Oligomerization Domains Are Almost Ubiquitous in the Collagen Superfamily, *J. Biol. Chem. 278*, 42200−42207.
16. 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.
17. Dauber-Osguthorpe, P., Maunder, C. M., and Osguthorpe, D. J. (1996) Molecular dynamics: deciphering the data, *J. Comput.-Aided Mol. Des. 10*, 177−185.
18. Bornstein, P., and Traub, W. (1979) The chemistry and biology of collagen, in *The Proteins* (Neurath, H., and Hill, R., Eds.) Vol. 4, pp 411−632, Academic Press, New York.
19. Sims, T. J., Miles, C. A., Bailey, A. J., and Camacho, N. P. (2003) Properties of collagen in OIM mouse tissues, *Connect. Tissue Res. 44* (Suppl. 1), 202−205.
20. Cabral, W. A., Fertala, A., Green, L. K., Korkko, J., Forlino, A., and Marini, J. C. (2002) Procollagen with skipping of α1(I) exon 41 has lower binding affinity for α1(I) C-telopeptide, impaired in vitro fibrillogenesis, and altered fibril morphology, *J. Biol. Chem. 277*, 4215−4222.
21. Prockop, D. J., and Fertala, A. (1998) Inhibition of the self-assembly of collagen I into fibrils with synthetic peptides. Demonstration that assembly is driven by specific binding sites on the monomers, *J. Biol. Chem. 273*, 15598−15604.
22. Hulmes, D. J., and Miller, A. (1979) Quasi-hexagonal molecular packing in collagen fibrils, *Nature 282*, 878−880.
23. Knott, L., and Bailey, A. (1998) Collagen cross-links in mineralizing tissues: a review of their chemistry, function, and clinical relevance, *Bone 22* (3), 181−187.
24. Bank, R. A., Robins, S. P., Wijmenga, C., Breslau-Siderius, L. J., Bardoel, A. F., van der Sluijs, H. A., Pruijs, H. E., and TeKoppele, J. M. (1999) Defective collagen cross-linking in bone, but not in ligament or cartilage, in Bruck syndrome: indications for a bone-specific telopeptide lysyl hydroxylase on chromosome 17, *Proc. Natl. Acad. Sci. U.S.A. 96* (3), 1054−1058.

BI048304B